Prediction Models for Prolonged Intensive Care Unit Stay After Cardiac Surgery

Systematic Review and Validation Study

Roelof G.A. Ettema, MSc; Linda M. Peelen, PhD; Marieke J. Schuurmans, PhD, RN; Arno P. Nierich, MD, PhD; Cor J. Kalkman, MD, PhD; Karel G.M. Moons, PhD

Background—Several models have been developed to predict prolonged stay in the intensive care unit (ICU) after cardiac surgery. However, no extensive quantitative validation of these models has yet been conducted. This study sought to identify and validate existing prediction models for prolonged ICU length of stay after cardiac surgery.

Methods and Results—After a systematic review of the literature, the identified models were applied on a large registry database comprising 11 395 cardiac surgical interventions. The probabilities of prolonged ICU length of stay based on the models were compared with the actual outcome to assess the discrimination and calibration performance of the models. Literature review identified 20 models, of which 14 could be included. Of the 6 models for the general cardiac surgery population, the Parsonnet model showed the best discrimination (area under the receiver operating characteristic curve=0.75 [95% confidence interval, 0.73 to 0.76]), followed by the European system for cardiac operative risk evaluation (EuroSCORE) (0.71 [0.70 to 0.72]) and a model by Huijskes and colleagues (0.71 [0.70 to 0.73]). Most of the models showed good calibration.

Conclusions—In this validation of prediction models for prolonged ICU length of stay, 2 widely implemented models (Parsonnet, EuroSCORE), although originally designed for prediction of mortality, were superior in identifying patients with prolonged ICU length of stay. **(*Circulation*. 2010;122:682-689.)**

Key Words: cardiovascular diseases ■ complications ■ epidemiology ■ risk factors ■ surgery

In the past decades, mortality during or shortly after cardiac surgery has decreased.[1] However, morbidity has increased,[2] mainly because cardiac surgery is increasingly utilized in older and more vulnerable patients. This often results in more complications after surgery and potential reduction in quality of life.[3–5] One method of assessing complications occurring directly after cardiac surgery is a prolonged stay in the intensive care unit (ICU).[6–9] Prolonged ICU stay also leads to incremental use of resources. In practice, prediction models are being used for efficient use of ICU resources. Patients with a low risk of complications are being scheduled for surgery before patients with a high risk.[5–13] Various prediction models have been developed to preoperatively identify patients with an increased risk for postoperative complications and prolonged ICU stay.[12–28] Interestingly, all of these prediction models were derived from samples including different patients, as reflected by the different distributions of patient and outcome characteristics. Hence, which model should be preferred in which situation is still unclear. Recently, in a qualitative review, Messaoudi and colleagues[14] reviewed 13 of these prediction models by comparing their published prognostic values for predicting ICU stay. They found that the 13 different prediction models indeed used different definitions of prolonged ICU stay and different definitions of predictors.

Clinical Perspective on p 689

Even though it is widely accepted that no prediction model should be applied in practice before being formally validated on its predictive accuracy in new patients,[29–31] no study has previously performed a formal, quantitative (external) validation of these prediction models in an independent patient population. Therefore, we first conducted a systematic review to identify all existing prediction models for prolonged ICU length of stay (PICULOS) after cardiac surgery. Subsequently, we validated the performance of the identified models in a large independent cohort of cardiac surgery patients.

Methods

Systematic Literature Review

In February 2008, the MEDLINE and PreMEDLINE databases were searched for studies on prediction models for PICULOS after cardiac

Received November 27, 2009; accepted June 4, 2010.

From the University of Applied Science Utrecht, Faculty of Health Care, Utrecht (R.G.A.E., M.J.S.); Julius Center for Health Sciences and Primary Care (L.M.P., K.G.M.M.) and Division of Perioperative Care and Emergency Aid (C.J.K.), Department of Rehabilitation, Nursing Science & Sports (M.J.S.), University Medical Center Utrecht, Utrecht; and Department of Anesthesiology and Intensive Care, Isala Clinics, Zwolle (A.P.N.), Netherlands.

The online-only Data Supplement is available with this article at http://circ.ahajournals.org/cgi/content/full/CIRCULATIONAHA.109.926808/DC1.

Correspondence to Roelof G.A. Ettema, University of Applied Science Utrecht, Faculty of Health Care, Bolognalaan 101 3584 CJ Utrecht, Netherlands. E-mail Roelof.Ettema@hu.nl

Circulation is available at http://circ.ahajournals.org DOI: 10.1161/CIRCULATIONAHA.109.926808

surgery that were published after 1980. The precise search query is presented in Appendix I in the online-only Data Supplement.

The retrieved articles were reviewed by 2 reviewers (R.G.A.E. and L.M.P.) and retained when they presented a formally developed prediction model. There is no consensus on the exact definition of PICULOS.[14] To relate to clinical practice,[2,11,13,15–28] we further restricted our analysis to prediction models that used a threshold for PICULOS within the bounds of 24 to 72 hours.

Application of the Models to an Independent Cohort

The validation of the retrieved models was then performed on a large cohort of cardiac surgery patients who underwent surgery between January 1, 2000, and July 31, 2008, at the Isala Clinics, Zwolle, Netherlands (1400 cardiac surgery procedures per year). The data had been collected prospectively as part of a continuous data registry for the national cardiac surgery patient registration. All patients provided informed consent to use the data for research. Patients' identifying information was removed before the analysis.

When the original articles did not provide sufficient information on the included predictors or regression coefficients (log odds ratios) in the model, the authors were asked to personally provide this information. If the information obtained was insufficient to apply the model to our data, the study was excluded from the analysis.

To validate the performance of the retrieved models, we used the original formulas and applied them to our patients using their observed predictor values. This yielded a predicted probability of PICULOS for each patient based on each model. To do this, we first matched the predictors in each prediction model to the variables in our data set. When a predictor was not available in our data set, we proceeded as follows. First, we sought to replace the variable with a proxy variable. Second, if a proxy was not available, we imputed the incidence or mean value reported in the literature for these predictors.[32–34] To prevent overimputation, this option was applied only when the weight of the predictor in the corresponding prediction model was relatively low compared with the other predictors in that model because it has a tempering effect on the predictive ability of the model. As a consequence, we only used this method for the predictors "family history" in the Parsonnet model[35,36] and "preoperative hemoglobin level" in the model of Huijskes et al.[17,37] If neither of these methods could be applied, the model was excluded from the analysis.[13,25]

Data Analysis

To analyze the performance of each prediction model, each patient's predicted probability of PICULOS in each model was compared with the observed outcome (ie, whether the patient had actually experienced PICULOS [yes/no]). To allow for a fair comparison of the models, a threshold for observed PICULOS had to be chosen. On the basis of the literature[15,17,21,22,25,26,28] and current clinical practice, we defined observed PICULOS as an ICU length of stay of >48 hours.

In comparing the performance of the models, we focused on discrimination and calibration. The discrimination performance of a model indicates the extent to which the model distinguishes between patients with and without prolonged ICU stay. The discrimination performance of the models was expressed by constructing receiver operating characteristic curves for each of the models and calculating the area under the curve (AUC) with a 95% confidence interval.[38] Theoretically, the AUC ranges from 0.5 (no predictive ability at all) to 1 (perfect predictive ability). In practice, however, the AUC can be well below the theoretical maximum of 1 even if the prediction model is perfectly calibrated, especially in complex diseases.[39]

The calibration performance of a model describes the extent to which the predicted probability of prolonged ICU stay reflects the true probability of prolonged ICU stay. The calibration of the models was judged by constructing calibration plots,[40] relating the predicted and observed probabilities. The calibration performance of a prediction model in an independent data set (external validation set) is commonly influenced by the incidence of the outcome in the validation set.

To allow for a fair comparison of the models, we adjusted the intercept of each model before applying it to the data, such that the mean predicted probability was equal to the observed outcome frequency.[34,41] Calibration plots were constructed subsequently. For each model, the *U* statistic (which compares the actual slope and intercept of the calibration plot to the ideal values of 1 and 0, respectively) was calculated and tested against a χ^2 distribution with 2 degrees of freedom.[33]

To further measure the accuracy of the models, we calculated the Yates slope (difference between the mean predicted probabilities for the patients with and without actual prolonged ICU stay) and the Brier scores (quadratic difference between predicted probability and actual outcome [0 or 1] for each patient) for each of the models.[42] All of these measures provide insight into the distance the model creates between the patients with and without a prolonged ICU stay.

Missing values occurred for the variables "gender" (0.05%), "myocardial infarction" (0.14%), "serum creatinine" (2.86%), "smoking" (0.13%), "height" and "weight" (both 45% of cases), "New York Heart Association classes" (0.92%), and the outcome variable "ICU length of stay" (1.71%). Missing values were substituted by means of single regression and weighted mean imputation, both of which are widely known methods for the substitution of missing values to reduce bias and increase statistical power.[32] Two-sided statistical tests were conducted with a significance level of 0.05. The statistical package R (version 2.10.1 [2009-12-14], The R Foundation for Statistical Computing, Vienna, Austria) was used for statistical analysis.

Results

Systematic Literature Review

Figure 1 shows the flow chart of the systematic literature review. From the 56 articles that matched the initial search query, 25 articles described 20 different prediction models.* Two models were excluded because they used a threshold of >72 hours to define prolonged ICU stay.[8,12] Additional information on intercepts, coefficients, and definitions of predictors in the models was requested from the authors for 7 models.[11,15,16,24,25,27,35] Two authors responded with the requested information,[11,16] 2 authors responded but were not able to provide the requested information,[15,27] and 4 authors did not respond. Three models were excluded because necessary information in regard to the definitions of the variables used was missing,[10,15,25] and 1 model was excluded because no adequate information was available in the database.[13] Finally, 14 prediction models could be included in our validation study.

Six of these 14 models were developed for patients undergoing cardiac surgery in general,[17,21,26,27,35,43,44] whereas the 8 other models focused on patients undergoing isolated coronary artery bypass grafting (CABG) surgery.[2,16,18,20,23,24,27] Two of the 14 prediction models, the Parsonnet model[35] and the European system for cardiac operative risk evaluation (EuroSCORE),[43,44] were originally designed for the prediction of mortality after cardiac surgery but have been used and validated for prolonged ICU stay.[7,8,11,21,28] Therefore, these models were also included in our study.

Table 1 describes the general characteristics of the 14 selected prediction models. Appendix II in the online-only Data Supplement provides a more extensive overview of the characteristics of the prediction models according to the framework established by Laupacis and colleagues.[45]

Predictive Performance

Table 2 describes the baseline characteristics of the patients in our cohort. We tested the prediction models in our cohort on the type of patients for which they were developed;

*References 2, 8, 11–13, 15–28, 35, 43, 44.

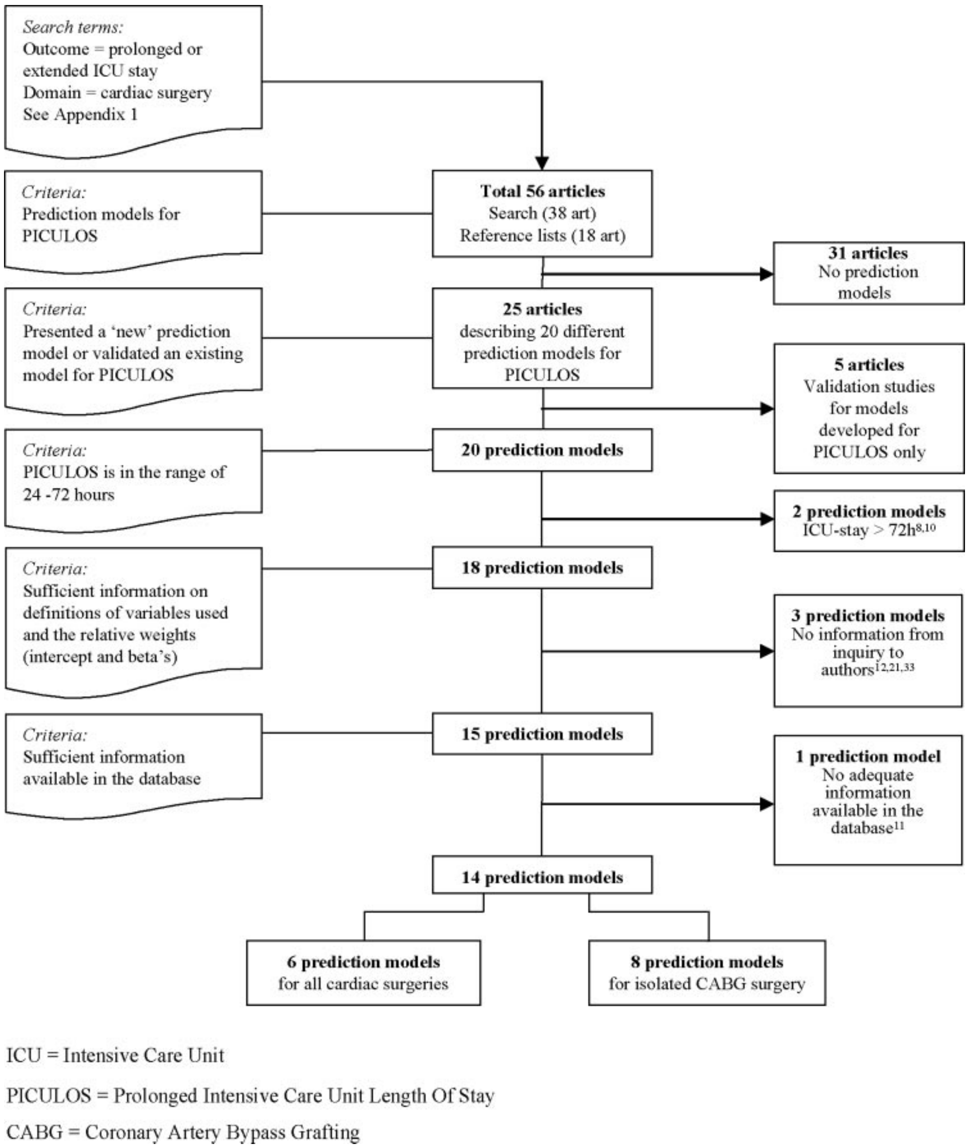

Figure 1. Flowchart of the systematic review of prediction models for prolonged ICU stay after cardiac surgery.

prediction models developed for cardiac surgery in general were evaluated on all patients (n=11 395), and prediction models developed for isolated CABG patients were evaluated on patients who underwent isolated CABG (n=6463) only.

Figure 2 depicts the receiver operating characteristic curves for each of the models, and Table 3 depicts the accompanying statistics. Among models including all cardiac surgeries, the Parsonnet model[8,11,35] showed the best discrimination (AUC 0.75 [95% confidence interval, 0.73 to 0.76]), followed by the EuroSCORE[7,21,28,43,44] (0.71 [0.70 to 0.72]) and a model by Huijskes and colleagues[17] (0.71 [0.70 to 0.73]). Among the models specifically developed for patients undergoing isolated CABG, the models by Wong et al,[23] Ivanov et al,[20] and Tuman et al[27] showed the best discrimination, with AUCs of 0.68 (0.65 to 0.70), 0.67 (0.65 to 0.70), and 0.66 (0.64 to 0.68], respectively.

Figure 3A and 3B show calibration plots of the 2 best- and the 2 least-performing models after adjustment of the intercept of each model for all cardiac surgery patients and isolated CABG patients, respectively. For most of the models, the calibration line in the plot closely followed the ideal calibration line, except for the models of Wong et al[23] and Abrahamyan et al.[16] The 6 models for the general cardiac surgery population had low *P* values for the *U* statistic (Table 3), indicating that the 6 models do not provide accurate probabilities. For the isolated CABG surgery patients, only the models of Tuman et al[27] and Christakis I (containing only preoperative predictors)[24] had nonsignificant *P* values.

Discussion

We conducted a systematic review and validated the performance of 14 retrieved prediction models to identify patients with prolonged ICU stay after cardiac surgery, using a large cohort of cardiac surgery patients. In this first quantitative comparison of all prediction models to identify patients who are likely to have a prolonged ICU stay, the Parsonnet model and the EuroSCORE show the best performance in terms of discrimination, accuracy, and calibration. Although both

Table 1. General Characteristics of the Studied Prediction Models

	Year of Publication	Period of Data Collection	Region (No. of Centers)	No. of Subjects in Development Set	Original Outcome	No. of Predictors	AUC in Initial Publication	*P*, HL Goodness-of-Fit Test in Initial Publication
Cardiac surgery								
Parsonnet[35]	1989	1982–1987	US (1)	3500	>24 h†	17	0.7†	NR
Tuman[27]	1992	NR	US (1)	3156	‡	16	NR	NR
Tu[26]	1994	1990–1991	Canada (1)	713	>2 d	10	0.69	0.24
EuroSCORE[43,44]	1999	1995	Europe (132)	13 302	>2 d§	20	0.78§	0.4§
							0.76 and 0.79§‖	
Pitkänen[21]	2000	1992–1996	Finland (1)	3061	>2 d	12	0.75 and 0.81‖	0.4 and 0.48‖
Huijskes[17]	2003	1997–2001	Netherlands (1)	4843	>2 d	14	0.79 and 0.78¶	0.63 and 0.36¶
Isolated CABG surgery								
Tuman[27]	1992	NR	US (1)	3156	‡	11	NR	NR
Christakis[24]*	1996	1990–1992	Canada (1)	889	>3 d	4	NR	NR
Wong[23]	1999	1995	Canada (1)	885	>2 d	9	0.89 and 0.85#	NR
Ivanov[20]	2000	1993–1997	Canada (2)	5354	>2 d	17	0.71	0.51
Janssen[18]	2004	2000–2001	Netherlands (1)	888	≥3 d	6	NR	NR
Abrahamyan[16]	2006	2003	Armenia (1)	391	≥3 d	4	0.71	0.6
Ghotkar[2]	2006	1997–2002	England (1)	5168	>3 d	14	0.72 and 0.74¶	0.3 and 0.79¶

HL indicates Hosmer-Lemeshow; NR, not reported.

*Two models were provided: 1 model containing only preoperative predictors (Christakis I) and 1 model containing both preoperative and postoperative predictors (Christakis II).

†Originally developed for mortality, validated for PICULOS by Hsieh et al (2007)[8] and Lawrence et al (2000).[11]

‡Group of patients with a mean ICU stay of 2.5 (±0.4) days compared with group of patients with a mean ICU stay of 7.0 (±9.6) days.

§Originally developed for mortality, validated for PICULOS by Pinna Pintor et al (2003),[7] Pitkänen et al (2000),[21] and Nilsson et al (2004).[28]

‖Figures based on a retrospective data set and a prospective data set, respectively.

¶Figures based on a derivation set and a validation set, respectively.

#Figures based on a derivation set and bootstrap validation, respectively.

models were originally developed to predict mortality, we found that they are also superior in identifying patients with an increased risk of prolonged ICU stay. A major explanation lies in the fact that, in current practice, mortality has decreased but morbidity has increased.[1,2] Because of advances in perioperative care in cardiac surgery,[46] most of the patients who were likely to die in the era when the Parsonnet model and the EuroSCORE were developed will now survive, but they still have a higher probability of developing complications. This is also supported by Parolari and colleagues,[47] who noted a significant overestimation of mortality with the EuroSCORE. Because both models overestimate mortality in current practice, these models for mortality need to be corrected for improved level of care in the future.

In the systematic review, we found 20 prediction models for prolonged ICU stay, 14 of which we could include in our analysis. In accordance with Messaoudi et al,[14] we found considerable differences in the definitions of the predictors and outcomes. We chose to restrict our systematic review to prediction models that used a threshold for PICULOS within the bounds of 24 to 72 hours. Afterward, in our validation study, we used the threshold of 48 hours because this correlates best with clinical practice. To verify the extent to which this difference has influenced our findings, we repeated the validation analysis using threshold values of 24 and 72 hours. This did not influence the ranking of the models based on their performance.

Table 2. Baseline Characteristics of Patients in the Database

	All Cardiac Surgery	Isolated CABG
No. of cases, n (% of total)	11 395 (100)	6463 (56.7)
PICULOS,* n (incidence in %)	1842 (16.1)	566 (8.8)
Female gender, n (%)	3397 (29.8)	1564 (24.2)
Age, y, median (Q 1, Q 3)	67.8 (59.6, 74.2)	66.8 (59.1, 73.1)
ICU days, median (Q 1, Q 3)	0.92 (0.8, 1.2)	0.91 (0.8, 1.0)

Q indicates quartile.

*Defined as ICU length of stay >48 hours.

Substantial differences between the models were also found in the sizes of the databases used to develop the prediction models and in the number of predictors in the models. Only 10 of the 14 models were initially validated, 9 of which used an independent validation set,† and 1 was validated by means of bootstrapping.[23] Prospective validation, however, was done for only 4 models.[21,35,44] In every case, the validation of the models was done in relatively small data sets (sizes ranging from 394[27] to 2439[17]). Only Parsonnet's model,[8,11] the EuroSCORE,[7,21,28] and Tu's model[8] were validated by other authors in a different geographic region. Because of all of these differences, the results of these original analyses are difficult to compare.

Our analysis is the first extensive quantitative validation of existing models for prolonged ICU stay after cardiac surgery in a large data set including >11 000 patients. All models were validated on the same data set, which allows for a proper comparison of the performance of the models.

†References 2, 17, 20, 21, 26, 27, 35, 43, 44.

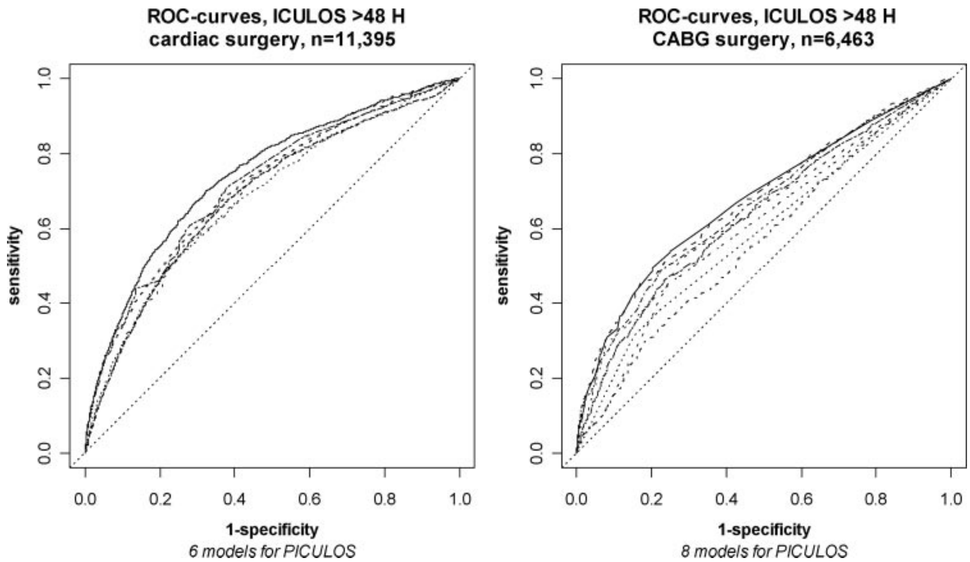

Figure 2. Receiver operating characteristic (ROC) curves for all 14 prediction models. The diagonal line represents zero discriminative value and corresponds to an AUC of 0.50. ICULOS indicates ICU length of stay.

To determine the calibration of the models, we made calibration plots and calculated the *U* statistic and the Hosmer-Lemeshow statistic. At first sight, these approaches gave contradictory results. In most of the models, the Hosmer-Lemeshow and *U* statistics had a *P* value <0.05, suggesting that the predictions based on the model deviated significantly from the observed data. In contrast, the calibration lines in the plots were very close to the 45° line, suggesting near-perfect calibration. To gain insight into the cause of these large statistics, we furthermore calculated the *t* values for the slopes of the models. This revealed that the slopes of the models in this data set deviate significantly from the ideal slope of 1. This would explain the large χ^2 values even after recalibration by adjusting the intercepts only. Whereas calibration statistics are merely summary measures, calibration plots directly reveal the variation of the performance of the model over the entire range of probabilities.[48]

Table 3. Predictive Performance of Prediction Models in the Study Cohort

	No. of Predictors in Model	Yates Slope*	Brier Score*	Brier Scaled*	AUC C Statistic*	*U* Statistic P (χ^2)* (Recalibrated Models)	Mean Predicted Risk, ICU LOS >48 h* (Recalibrated Models)	Mean Predicted Risk, ICU LOS >48 h* (Not Recalibrated Models)
All cardiac (n=11 395)							Incidence ICU LOS >48 h=0.162	
Parsonnet	17	0.157	0.122	0.065	0.75 (0.73–0.76)	<0.000 (100.61)	0.162	0.066
Tuman	16	0.079	0.128	0.064	0.67 (0.66–0.69)	<0.000 (15.28)	0.162	NA†
Tu	10	0.099	0.129	0.064	0.69 (0.68–0.71)	<0.000 (154.88)	0.162	0.357
EuroSCORE	20	0.149	0.126	0.064	0.71 (0.70–0.72)	<0.000 (397.58)	0.162	0.877
Pitkänen	12	0.096	0.130	0.064	0.69 (0.67–0.70)	<0.000 (226.46)	0.162	0.206
Huijskes	14	0.155	0.127	0.064	0.71 (0.70–0.73)	<0.000 (305.96)	0.162	0.049
Isolated CABG (n=6463)							Incidence ICU LOS >48 h=0.088	
Tuman	11	0.046	0.076	0.115	0.66 (0.64–0.68)	0.383 (1.92)	0.097	NA†
Christakis, preoperative predictors	4	0.014	0.080	0.115	0.59 (0.56–0.61)	0.116 (4.31)	0.085	NA†
Christakis, preoperative and postoperative predictors	4	0.050	0.084	0.114	0.62 (0.60–0.64)	<0.000 (205.53)	0.095	NA†
Wong	9	0.135	0.076	0.114	0.68 (0.65–0.70)	<0.000 (474.92)	0.103	NA†
Ivanov	17	0.082	0.080	0.115	0.67 (0.65–0.70)	<0.000 (29.72)	0.090	0.299
Janssen	6	0.048	0.079	0.115	0.63 (0.60–0.65)	<0.000 (78.33)	0.098	0.167
Abrahamyan	4	0.031	0.087	0.114	0.57 (0.54–0.59)	<0.000 (433.61)	0.089	0.767
Ghotkar	14	0.051	0.081	0.115	0.64 (0.60–0.66)	<0.000 (83.61)	0.086	0.143

LOS indicates length of stay; NA, not applicable.

*All statistics are scaled from 0 to 1. Higher Yates slope, as well as lower Brier Scores and higher Brier Scaled and higher discrimination C statistics and nonsignificant *P* values of the calibration *U* statistic, represent better performance.

†The mean predicted risk for these (not calibrated) models could not be calculated because the original intercepts were not provided for these models.

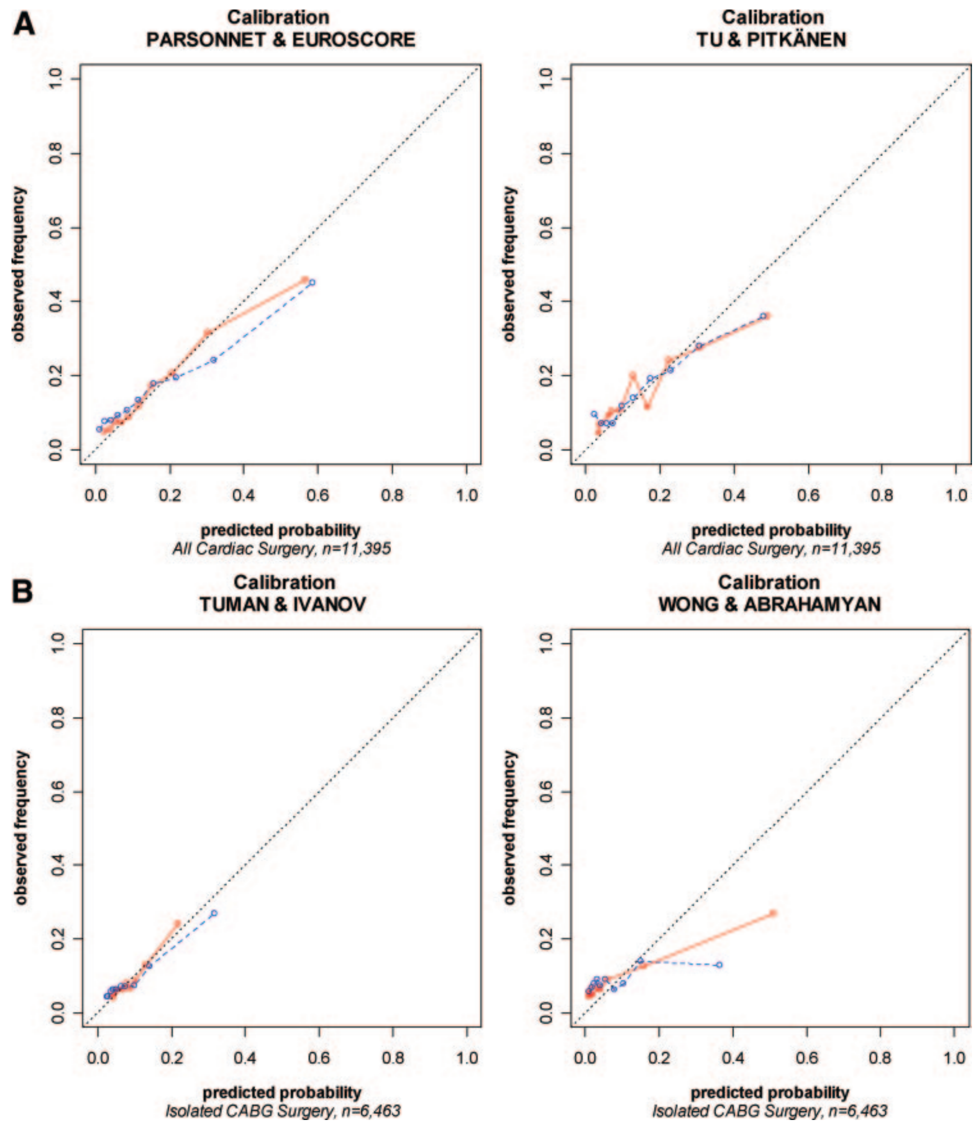

Figure 3. A, Calibration plots for models for all cardiac surgery. One plot for the 2 best-performing models (Parsonnet [solid red line] and EuroSCORE [dashed blue line]) and 1 plot for the 2 least-performing models (Tu [solid red line] and Pitkänen [dashed blue line]) are shown. The dotted line represents ideal calibration (with intercept 0 and regression coefficient 1); n=11 395. B, Calibration plots for models for isolated CABG surgery. One plot for the 2 best-performing models (Tuman [solid red line] and Ivanov [dashed blue line]) and 1 plot for the 2 least-performing models (Wong [solid red line] and Abrahamyan [dashed blue line]) are shown The dotted line represents ideal calibration (with intercept 0 and regression coefficient 1); n=6463.

Table 3 also shows the importance of recalibrating a model by adjusting the intercept[34,41] before calibration of the model is assessed. The mean predicted risks of the original models do not even approach the observed outcome frequency, whereas after recalibration this problem is solved. This allows for a more fair comparison of the models and a better performance when the models are applied in daily practice.

To determine the discrimination performance of the models, we calculated the AUCs. The 6 models for the general cardiac surgery population yielded AUCs ranging from 0.68 to 0.74. In the models specifically developed for patients with isolated CABG surgery, substantially lower AUCs (0.56 to 0.67) were found. In general, values for the AUC <0.70 indicate that use of the model in clinical practice should be done with caution[49] because the theoretical maximum value of the AUC is 1.0. However, it is also known that in practice this maximum depends not only on the model but also on characteristics of the data.[39] To allow for better interpretation of our findings and provide a "benchmark value," we fitted 2 reference models on the data (1 for all patients, 1 for patients undergoing isolated CABG only), which yielded AUCs of 0.80 and 0.73, respectively. These models are likely to be overfit but give a reference value for interpretation of the AUCs of the prediction models found in the literature.

We also found a considerable difference in AUC between the models for all patients and the models predicting prolonged ICU stay after isolated CABG procedures. To investigate whether this was due to the models or due to the differences in population characteristics (isolated CABG patients versus all patients), the 6 models for the general population were also applied to the isolated CABG patients only, resulting again in AUCs varying from 0.55 to 0.69. These AUCs are comparable to the AUCs of the models specially developed for isolated CABG surgery patients. This suggests that it is more complicated to predict prolonged ICU stay in isolated CABG surgery patients than in the cardiac surgery population as a whole. The Parsonnet model and the EuroSCORE showed the best discrimination (0.69 and 0.68, respectively) in the CABG surgery population. The Parsonnet model performed even better than the best-performing model (Wong, 0.68) that was specially developed for CABG surgery patients only.

Limitations

Obviously, prolonged ICU stay is intrinsically a continuous variable (length of stay). Accordingly, as with most continuous variables in medicine, one would rather not dichotomize[14] but would rather predict the original length of stay value itself. However, all published models used as outcome dichotomized prolonged stay (length of stay with some threshold value), and our purpose was to validate these models as published.

We made use of a prospective continuous data registry that includes all patients who underwent surgery and systematically recorded a large amount of information on preoperative, perioperative, and postoperative characteristics. A disadvantage of using registry data is that not all predictors of the models are available in the registry with exactly the same definition as used to develop these models. We have solved this problem in part by using proxy variables and by replacing missing variables with the incidence or mean of the predictor based on the literature. When too many concessions had to be made before the model could be applied to our data, we excluded the prediction model from this validation study.[8,10,12,13,15,25] Therefore, we do not think that the use of registry data has significantly influenced our conclusions. On the contrary, by using registry data we validated the performance of the models in daily clinical practice, which was specifically the aim of our study.

For most of the variables in the data set, the percentage of missing data was small. For height and weight, however, data were missing in 45% of the cases. Deleting 45% of the patient records (doing a complete case analysis) is widely known to yield biased results.[32] We thus applied the best available methods to properly deal with these missing data and minimize this bias and explicitly chose to impute the data by fitting a model.[32,50] With a percentage as high as 45% for missing data for 2 variables, theoretically multiple imputation is to be preferred over single imputation. However, in the context of multiple imputation, the manner in which to estimate the standard errors of part of the performance measures we used in this study is not straightforward. We have performed multiple imputation as a sensitivity analysis and found similar results for the point estimates, indicating that the numbers presented in this article are not influenced by the choice of the imputation strategy. We realize that we made use of data from a single center over a longer time period, which must be taken into account when our findings are generalized.

Conclusions

This extensive quantitative validation study demonstrates that the widely implemented Parsonnet and EuroSCORE models are superior to other models in predicting prolonged ICU stay after cardiac surgery. In current daily practice, Parsonnet's model and the EuroSCORE are widely implemented for the prediction of mortality risk. This allows for the relatively straightforward application of our findings in clinical practice. The predictions that have already been made for mortality can also be used to identify patients with a high probability of prolonged ICU stay. This knowledge, when available before surgery, can be used for timely planning of postoperative care and ICU management.

Acknowledgments

Professor Moons gratefully acknowledges The Netherlands Organization for Scientific Research for their financial support (project nr. 016.046.360 and 912.08.004).

Disclosures

None.

References

1. Northrup WF, Emery RW, Nicoloff DM, Lillehei TJ, Holter AR, Blake DP. Opposite trends in coronary artery and valve surgery in a large multisurgeon practice, 1979–1999. *Ann Thorac Surg*. 2004;77:488–495.
2. 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. *Eur J Cardiothorac Surg*. 2006;31:1–14.
3. Litovski DS, Dacey LJ, Baribeau YR, Leavitt BJ, Clough R, Cochran RP, Quinn R, Sisto DA, Charlesworth DC, Malenka DJ, MacKenzie TA, Olmstead EM, Ross CS, O'Connor GT. Long-term survival of the very elderly undergoing coronary artery bypass grafting. *Ann Thorac Surg*. 2008;85:1233–123.
4. Pätilä T, Kukkonen S, Vento A, Pettilä V, Suojaranta-Ylinen R. Relation of the sequential organ failure assessment score to morbidity and mortality after cardiac surgery. *Ann Thorac Surg*. 2006;82:2072–2079.
5. Stoica SC, Sharples LD, Ahmed I, Roques F, Large SR, Nashef SAM. Preoperative risk prediction and intraoperative events in cardiac surgery. *Eur J Cardiothorac Surg*. 2002;21:41–46.
6. Bucerius J, Gummert JF, Walther T, Doll N, Falk V, Schmitt DV, Mohr FW. Predictors of prolonged ICU stay after on-pump versus off-pump coronary artery bypass grafting. *Intensive Care Med*. 2004;30:88–95.
7. Pinna Pintor P, Bobbio M, Colangelo S, Veglia F, Marras R, Diena M. Can EuroSCORE predict direct costs of cardiac surgery? *Eur J Cardiothorac Surg*. 2003;23:595–598.
8. Hsieh CH, Peng SK, Tsai TC, Shih YR, Peng SY. Prediction for major adverse outcomes in cardiac surgery: comparison of three prediction models. *J Formos Med Assoc*. 2007;106:759–767.
9. Scott BH, Seifert FC, Grimson R, Glass PSA. Octogenarians undergoing coronary artery bypass graft surgery: resource utilization, postoperative mortality, and morbidity. *J Cardiothorac Vasc Anesth*. 2005;19:583–588.
10. Salamonsen RS, Bailey MJ, Salamonsen RF. Can ICU stay be predicted accurately enough for fast-tracking cardiac surgical patients? *Crit Care Resusc*. 2008;10:194–201.
11. Lawrence DR, Valencia O, Smith EEJ, Murday A, Treasure T. Parsonnet score is a good predictor of the duration of intensive care unit stay following cardiac surgery. *Heart*. 2000;83:429–432.
12. 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;21:146–150.
13. Hein OV, Birnbaum J, Wernecke K, England M, Konertz W, Spies C. Prolonged intensive care unit stay in cardiac surgery: risk factors and long-term survival. *Ann Thorac Surg*. 2006;81:880–885.
14. Messaoudi N, De Cocker J, Stockman B, Bossaert LL, Rodrigus IER. Prediction of prolonged length of stay in the intensive care unit after cardiac surgery: the need for a multi-institutional risk scoring system. *J Card Surg*. 2009;24:127–133.
15. 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.
16. Abrahamyan L, Demirchyan A, Thompson ME, Hovaguimian H. Determinants of morbidity and intensive care unit stay after coronary surgery. *Asian Cardiovasc Thorac Ann*. 2006;14:114–118.
17. Huijskes RV, Rosseel PM, Tijssen JG. Outcome prediction in coronary artery bypass grafting and valve surgery in the Netherlands:

development of the Amphiascore and its comparison with the EuroSCORE. *Eur J Cardiothorac Surg*. 2003;24:741–749.
18. 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;25:203–207.
19. Geissler HJ, Hözl P, Marohl S, Kuhn-Régnier F, Mehlhorn U, Südkamp M, Rainer de Vivie E. Risk stratification in heart surgery: comparison of six score systems. *Eur J Cardiothorac Surg*. 2000;17:400–406.
20. Ivanov J, Borger MA, David TE, Cohen G, Walton N, Naylor CD. Predictive accuracy study: comparing a statistical model to clinicians' estimates of outcomes after coronary bypass surgery. *Ann Thorac Surg*. 2000;70:162–168.
21. 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;18:703–710.
22. Ivanov J, Tu JV, Naylor DC. Ready-made, recalibrated, or remodeled? Issues in the use of risk indexes for assessing mortality after coronary artery bypass graft surgery. *Circulation*. 1999;99:2098–2104.
23. Wong DT, Cheng DCH, Kustra R, Tibshirani R, Karski J, Carroll-Munro J, Sandler A. Risk factors of delayed extubation, prolonged length of stay in the intensive care unit, and mortality in patients undergoing coronary artery bypass graft with fast-track cardiac anesthesia: a new cardiac risk score. *Anesthesiology*. 1999;91:936–944.
24. Christakis GT, Fremes SE, Naylor CD, Chen E, Rao V, Goldman BS. Impact of preoperative risk and perioperative morbidity on ICU stay following coronary bypass surgery. *Cardiovasc Surg*. 1996;4:29–35.
25. Michalopoulos A, Tzelepis G, Pavlides G, Kriaras J, Dafni U, Geroulanos S. Determinants of duration of ICU stay after coronary artery bypass graft surgery. *Br J Anaesth*. 1996;77:208–212.
26. Tu JV, Mazer CD, Levinton C, Armstrong PW, Naylor CD. A predictive index for length of stay in the intensive care unit following cardiac surgery. *Can Med Assoc J*. 1994;151:177–185.
27. Tuman KJ, McCarthy RJ, March RJ, Najafi H, Ivankovich AD. Morbidity and duration of ICU stay after cardiac surgery: a model for preoperative risk assessment. *Chest*. 1992;102:36–44.
28. Nilsson J, Algotsson L, Höglund P, Lührs C, Brandt J. EuroSCORE predicts intensive care unit stay and costs of open heart surgery. *Ann Thorac Surg*. 2004;78:1528–1535.
29. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med*. 1999;130:515–524.
30. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med*. 2006;144:201–209.
31. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:b375.
32. Janssen KJ, Donders AR, Harrell FE Jr, Vergouwe Y, Chen Q, Grobbee DE, Moons KG. Missing covariate data in medical research: to impute is better than to ignore. *J Clin Epidemiol*. 2010;63:721–727.
33. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic model issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15:361–387.
34. Steyerberg EW, Eijkemans MJC, Van Houwelingen JC, Lee KJ, Habbema JDF. Prognostic models based on literature and individual patient data logistic regression analysis. *Stat Med*. 2000;19:141–160.
35. Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart disease. *Circulation*. 1989;79(suppl 1):I-3–I-12.
36. Nasir K, Budoff MJ, Wong ND, Scheuner M, Herrington D, Arnett DK, Szklo M, Greenland P, Blumenthal RS. Family history of premature coronary heart disease and coronary artery calcification: Multi-Ethnic Study of Atherosclerosis (MESA). *Circulation*. 2007;116:619–626.
37. Ouattara A, Niculescu M, Boccara G, Landi M, Vaissier E, Léger P, Riou B, Gandjbakch I, Coriat P. Identification of risk factors for allogenic transfusion in cardiac surgery from an observational study. *Ann Fr Anesth Reanim*. 2003;22:278–283.
38. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*. 1982;143:29–36.
39. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*. 2007;115:928–935.
40. Spiegelhalter DJ. Probabilistic prediction in patient management and clinical trials. *Stat Med*. 1986;5:421–433.
41. Janssen KJM, Moons KGM, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating a clinical prediction model improved the performance in new patients. *J Clin Epidemiol*. 2008;61:76–86.
42. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, Pencina MJ, Kattane MW. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21:128–138.
43. Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). *Eur J Cardiothorac Surg*. 1999;16:9–13.
44. Roques F, Nashef SAM, Michel P, Gauducheau E, de Vincentiis C, Baudet E, Cortina J, David M, Faichney A, Gabrielle F, Gams E, Harjula A, Jones MT, Pinna Pintor P, Salamon R, Thulin L. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. *Eur J Cardiothorac Surg*. 1999;15:816–823.
45. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules: a review and suggested modifications of methodological standards. *JAMA*. 1997;277:488–494.
46. Svircevic V, Nierich AP, Moons KG, Brandon Bravo Bruinsma GJ, Kalkman CJ, van Dijk D. Fast-track anesthesia and cardiac surgery: a retrospective cohort study of 7989 patients. *Anesth Analg*. 2009;108:727–733.
47. Parolari A, Pesce LL, Trezzi M, Loardi C, Kassem S, Brambillasca C, Miguel B, Tremoli E, Biglioli P, Alamanni F. Performance of EuroSCORE in CABG and off-pump coronary artery bypass grafting: single institution experience and meta-analysis. *Eur Heart J*. 2009;30:297–304.
48. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation and Updating*. New York, NY: Springer; 2009.
49. Harrell FE Jr. *Regression Modelling Strategies*. New York, NY: Springer; 2001.
50. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*. 2006;59:1087–1091.

CLINICAL PERSPECTIVE

Prolonged intensive care unit (ICU) stay after cardiac surgery leads to potential reduction in quality of life and incremental use of resources. For efficient use of ICU resources and to schedule patients with a low risk of postoperative complications before patients with a higher risk, preoperative estimation of the risk of prolonged ICU stay is necessary. Various prediction models have been developed to preoperatively identify patients with an increased risk for prolonged ICU stay. It is widely accepted that no prediction model should be applied in practice before being formally validated in new patients. In the domain of prolonged ICU stay after cardiac surgery, however, no study has thus far conducted such a formal validation and comparison study. The present analysis is the first extensive quantitative validation of existing models for prolonged ICU stay after cardiac surgery in a large data set of 11 395 patients. The results show that the Parsonnet model and the European system for cardiac operative risk evaluation (EuroSCORE) have the overall best performance. Although both models were originally developed to predict mortality, they are also superior in identifying patients with an increased risk of prolonged ICU stay. Because in current daily practice both models are widely implemented for the estimation of mortality risk, this allows for a relatively straightforward application of our findings in clinical practice. The risk stratification for mortality based on these models can also be used to identify patients with an increased risk of prolonged ICU stay, which is useful for timely planning of postoperative care and ICU management.

Circulation

American Heart Association®

Prediction Models for Prolonged Intensive Care Unit Stay After Cardiac Surgery: Systematic Review and Validation Study

Roelof G.A. Ettema, Linda M. Peelen, Marieke J. Schuurmans, Arno P. Nierich, Cor J. Kalkman and Karel G.M. Moons

Circulation. 2010;122:682-689; originally published online August 2, 2010;
doi: 10.1161/CIRCULATIONAHA.109.926808
Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 0009-7322. Online ISSN: 1524-4539

SUPPLEMENTAL MATERIAL

MANUSCRIPT ID:
CIRCULATIONAHA/2009/926808/R2

1) *Query used for the systematic review* (Supplemental Methods)
2) *Table 1a.* Methodological features (according to Laupacis38) of the six prediction models for PICULOS after all cardiac surgeries (Supplemental Table)

   Table 1b. Methodological features (according to Laupacis38) of the eight prediction models for PICULOS after isolated CABG surgery (Supplemental Table)

APPENDIX 1

Query used for the systematic review

The MEDLINE and PreMEDLINE databases were searched for publications concerning prediction models for prolonged ICU stay after cardiac surgery, using the following query:

("Coronary Artery Bypass" OR "Valve surgery" OR "cardiac surgery" OR "cardiovascular surgery" OR "cardiac surgery procedure") AND (algorithm OR "multivariate analysis" OR "logistic model" OR "biological model" OR "statistical model" OR mathematics OR "regression analysis" OR "risk factor" OR "risk assessment" OR "predictive value" OR "Area Under Curve" OR "evaluation study" OR evaluation OR reproducibility OR prediction OR "prediction rule" OR predict OR prognosis OR "prognostic factor") AND (complication OR "adverse event" OR prolonged OR extended) AND stay AND ("intensive care unit" OR ICU).

APPENDIX 2

Table 1a. Methodological features (according to Laupacis[38]) of the six prediction models for PICULOS after *all cardiac surgeries*

	Characteristics of the six selected prediction models for prolonged ICU length of stay (PICULOS) after all cardiac surgeries						
	DESCRIPTION OF:	**Parsonnet**	**Tuman**	**Tu**	**EuroSCORE**	**Pitkänen**	**Huijskes**
1	**Reference(s)**	Parsonnet et al. 1989[36]	Tuman et al. 1992[35]	Tu et al. 1994[34]	Nashef et al. 1999[30]; Roques et al. 1999[31]	Pitkänen et al. 2000[27]	Huijskes et al. 2003[23]
2	**Outcome studied**						
	• Definition	Postoperative mortality	Difference between patients with and without morbidity *	PICULOS >2 days	Postoperative mortality	PICULOS >2 days	PICULOS >2 days
	• Blind assessment	√	√	√	√	√	√
3	**Predictors**						
	• Definition predictors	√	√	√	√	√	√
	• Number of predictors in model	17	16	10	20	12	14
	• Methods of data collection	Available dataset	Prospective collection	Prospective collection	Prospective collection	Prospective collection	Prospective collection
4	**Patient characteristics**						
	• Data collection time frame	1982 - 1987	n.r.	1990 - 1991	1995	1992 - 1996	1997 - 2001
	• Procedure types	Cardiac surgery	Cardiac surgery	Cardiac surgery	Cardiac surgery	Cardiac surgery	Cardiac surgery
5	**Study site**						
	• No. of centres	1	1	1	132	1	1
	• Region	USA	USA	Canada	Europe	Finland	Netherlands
	• No. of patients in derivation cohort	3,500	3,156	713	13,302	3,061	4,843
6	**Mathematical techniques**						
	• Handling of missing data	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.
	• Handling of dichotomous,	n.r.	n.r.	n.r.	√	√	√

Characteristics of the six selected prediction models for prolonged ICU length of stay (PICULOS) after all cardiac surgeries							
	DESCRIPTION OF:	**Parsonnet**	**Tuman**	**Tu**	**EuroSCORE**	**Pitkänen**	**Huijskes**
	categorical and continuous variables						
	• Univariable and multivariable Analysis	√	√	√	√	√	√
7	**Results of the model**						
	• AUC	n.r. ‡	n.r.	0.69	0.79 & 0.76†	0.75 & 0.81‖	0.79& 0.78†
	• Calibration plots	√	√	n.r.	n.r.	√	n.r.
	• P-value HL goodness-of-fit	n.r. §	n.r.	0.24	0.4 & 0.68†	0.4 & 0.48‖	0.63& 0.36†
8	**Likelihood of use in practice**						
	• Clinicians perceive items in model as appropriate	√	√	√	√	√	√
	• Risk score	√	n.r.	√	√	n.r.	√
	• Probability of the outcome	√	√	√	√	√	√
	• Model not limited to a risk score, but also suggests a course of action	n.r.	n.r.	√	√	n.r.	n.r.
9	**Previously validated in external cohort**						
	in the initial study						
	• No. of centres	2	1	1	132	1	1
	• Region	USA	USA	Canada	Europe	Finland	Netherlands
	• No. of patients in validation cohort	1,332	394	691	1,479	153 & 82 ‖	2,439
	in an additional study	*ForPICULOS:*			*For PICULOS:*		
	• Reference	Lawrence et al. 2000[13]	-	Tu et al. 1996[8]	a) Pitkänen et al. 2000[27] b) Pina Pintor et al. 2003[7] c) Nilsson et al. 2004[37]	-	-
	• No. of centres (region)	1 (England)	-	1 (Canada)	*a) 1 (Finland)* *b) 1 (Italy)*	-	-

	Characteristics of the six selected prediction models for prolonged ICU length of stay (PICULOS) after all cardiac surgeries						
	DESCRIPTION OF:	**Parsonnet**	**Tuman**	**Tu**	**EuroSCORE**	**Pitkänen**	**Huijskes**
	• No. of patients in the cohort	5,591	-	265	*c) 1 (Sweden)* *a) 4,592* *b) 3,404* *c) 488*	-	-
10	**Effects of clinical use measured**	√	n.r.	√	√	n.r.	n.r.

ICU = intensive care unit

n.r. = not reported

AUC = area under the ROC (receiver operating characteristic) curve

HL = Hosmer-Lemeshow (p-value of the Hosmer-Lemeshow goodness-of-fit statistic)

* Difference between two groups: group 1 without morbidity (ICU stay 2.5 ± 0.4 days) and group 2 with morbidity (ICU stay 7.0 ± 9.6 days)

† Figures based on a derivation set and a validation set, respectively

‡ Instead of an AUC for the discriminative ability of the model, a mean correlation (Spearman's rho) was calculated (0.99)

§ Instead of a p-value of the Hosmer-Lemeshow goodness-of-fit statistic for the calibrative ability of the model, a group correlation (Spearman's rho) was calculated (0.85)

|| Figures based on a retrospective dataset and a prospective dataset, respectively

Table 1b. Methodological features (according to Laupacis[38]) of the eight prediction models for PICULOS after *isolated CABG surgery*

	Characteristics of the eight selected prediction models for prolonged ICU length of stay (PICULOS) after isolated CABG surgery							
	DESCRIPTION OF:	**Tuman**	**Christakis**	**Wong**	**Ivanov**	**Janssen**	**Abrahamyan**	**Ghotkar**
1	**Reference(s)**	Tuman et al. 1992[35]	**2 models:** Christakis et al. 1996[32]	Wong et al. 1999[29]	Ivanov et al. 1999[28]; Ivanov et al. 2000[26]	Janssen et al. 2003[24]	Abrahamyan et al. 2006[22]	Ghotkar et al. 2006[2]
2	**Outcome studied**							
	• Definition	Difference between patients with and without morbidity *	PICULOS >3 days	PICULOS >2 days	PICULOS >2 days	PICULOS ≥3 days	PICULOS ≥3 days	PICULOS >3 days
	• Blind assessment	√	√	√	√	√	√	√
3	**Predictors**							

	Characteristics of the eight selected prediction models for prolonged ICU length of stay (PICULOS) after isolated CABG surgery							
	DESCRIPTION OF:	**Tuman**	**Christakis**	**Wong**	**Ivanov**	**Janssen**	**Abrahamyan**	**Ghotkar**
	• Definition predictors	√	√	√	√	√	√	-
	• Number of predictors in model	11	4	9	17	6	4	14
	• Methods of data collection	Prospec-tive	Prospec-tive	Prospec-tive	Prospec-tive	Prospec-tive	Prospec-tive	Prospec-tive
4	**Patient population**							
	• Data collection time frame	n.r.	1990 – 1992	1995	1993 – 2007	2000 – 2001	2003	1997 – 2002
	• Procedure type	CABG	CABG	CABG	CABG	CABG	CABG	CABG
5	**Study site**							
	• No. of centres	1	1	1	2	1	1	1
	• Region	USA	Canada	Canada	Canada	Nether-lands	Armenia	England
	• No. of patients in derivation cohort	2,366	889	885	5,354	888	391	5,168
6	**Mathematical techniques**							
	• Handling of missing data	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.
	• Handling of dichotomous, categorical and continuous variables	n.r.	√	√	n.r.	√	n.r.	√
	• Univariable and multivariable analysis	√	√	√	√	√	√	√
7	**Results of the model**							
	• AUC	n.r.	n.r.	n.r.	0.71	n.r.	0.71	0.72 & 0.7†
	• Calibration plots	√	n.r.	n.r.	√	n.r.	n.r.	√
	• P-value HL goodness-of-fit	n.r.	n.r.	n.r.	0.51	n.r.	0.6	0.3 & 0.79†
8	**Likelihood of use in practice**							
	• Clinicians perceive items in model as appropriate	√	√	√	√	√	√	√

Characteristics of the eight selected prediction models for prolonged ICU length of stay (PICULOS) after isolated CABG surgery

	DESCRIPTION OF:	Tuman	Christakis	Wong	Ivanov	Janssen	Abrahamyan	Ghotkar
	• Risk score	n.r.	n.r.	√	n.r.	n.r.	n.r.	√
	• Probability of the outcome	√	n.r.	√	n.r.	√	n.r.	√
	• Model not limited to a risk score alone, but also suggests a course of action	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.	n.r.
9	**Previously validated in external cohort**							
	In the initial study							
	• No. of centres	1	n.r.	-	2	n.r.	n.r.	1
	• Region	USA	n.r.	-	Canada	n.r.	n.r.	England
	• No. of patients in validation cohort	394	n.r.	Bootstrap	2,148	n.r.	n.r.	1,197
	In an additional study							
	• Reference	-	-	-	Ivanov et al. 2000[26]	-	-	-
	• No. of centres	-	-	-	1	-	-	-
	• Region	-	-	-	Canada	-	-	-
	• No. of patients in the cohort	-	-	-	1,904	-	-	-
10	**Effects of clinical use measured**	n.r.	n.r.	n.r.	√	n.r.	n.r.	n.r.

ICU = intensive care unit

CABG = coronary artery bypass grafting

AUC = area under the ROC (receiver operating characteristic) curve

HL = Hosmer-Lemeshow (p-value of the Hosmer-Lemeshow goodness-of-fit statistic, measure of calibrative ability.)

n.r. = not reported

* Difference between two groups: group 1 without morbidity (ICU stay 2.5 ± 0.4 days) and group 2 with morbidity (ICU stay 7.0 ± 9.6 days)

† Figures based on a derivation set and a validation set, respectively