ELSEVIER

European Journal of Cardio-thoracic Surgery 18 (2000) 703–710

EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY

www.elsevier.com/locate/ejcts

Intra-institutional prediction of outcome after cardiac surgery: comparison between a locally derived model and the EuroSCORE

Otto Pitkänen[a,*], Minna Niskanen[a], Sinikka Rehnberg[b], Mikko Hippeläinen[c], Markku Hynynen[a]

[a]*Department of Anaesthesiology and Intensive Care, Kuopio University Hospital, FIN-70210 Kuopio, Finland*
[b]*Department of Internal Medicine, Kuopio University Hospital, FIN-70210 Kuopio, Finland*
[c]*Department of Surgery, Kuopio University Hospital, FIN-70210 Kuopio, Finland*

Received 5 May 2000; received in revised form 9 August 2000; accepted 19 September 2000

Abstract

Objective: To construct models for predicting mortality, morbidity and length of intensive care unit (ICU) stay after cardiac surgery and to compare the performance of these models with that of the EuroSCORE in two independent validation databases. **Methods**: Clinical data on 4592 cardiac surgery patients operated between 1992 and 1996 were retrospectively collected. In order to derive predictive models and to validate them, the patient population was randomly divided into a derivation database ($n = 3061$) and a validation database ($n = 1531$). Variables that were significant in univariate analyses were entered into a backward stepwise logistic regression model. The outcome was defined as mortality within 30 days after surgery, predefined morbidity, and the length of ICU stay lasting > 2 days. In addition to the retrospective database, the models were validated also in a prospectively collected database of cardiac surgical patients operated in 1998–1999 ($n = 821$). The EuroSCORE was tested in two validation databases, i.e. the retrospective and prospective one. Hosmer–Lemeshow goodness-of-fit was used to study the calibration of the predictive models. Area under the receiver operating characteristic (ROC) curve was used to study the discrimination ability of the models. **Results**: The overall mortality in the retrospective and the prospective data was 2 and 1%, and morbidity 22 and 18%, respectively. The created predictive models fitted well in the validation databases. Our models and the EuroSCORE were equally good in discriminating patients. Thus, in the prospective validation database, the mean areas under the ROC curve for our models and for the EuroSCORE were similar, i.e. 0.84 and 0.77 for mortality, 0.74 and 0.74 for morbidity, and 0.81 and 0.79 for the length of intensive care unit stay lasting for 2 days or more, respectively. **Conclusions**: Our models and the EuroSCORE were equally good in discriminating the patients in respect to outcome. However, our model provided also well calibrated estimation of the probability of prolonged ICU stay for each patient. As was originally suggested, the EuroSCORE may be an appropriate tool in categorizing cardiac surgical patients into various subgroups in interinstitutional comparisons. Our models may have additive value especially in resource allocation and quality assurance purposes for local use.

Keywords: Cardiac surgery; Risk stratification; Mortality; Morbidity; Length of intensive care unit stay

1. Introduction

In adult cardiac surgery, there has been an increasing interest for prediction of operative mortality, morbidity and resource use after surgery [1–10]. The aims have been to create rules or models to identify patient characteristics associated with the probability of adverse outcomes and, thereby, to clarify criteria for patient selection and resource allocation. With these prediction models, the interest has also been to compare different institutions and surgeons in their performance by plotting observed patient outcome against predicted risk [11–13].

Most of the risk indexes are based on logistic regression analysis. They are either risk scores based on the odds ratios of variables in the logistic regression analysis or the indexes can also be risk-equations that produce a definite probability of mortality or morbidity [1–3,10,12]. Risk indexes are most valid in patient populations where the preoperative patient characteristics and treatment protocols are comparable with those of the original environments. That is why a model should not be used elsewhere as such, before its validity has been tested in the local patient material [8]. Recently, early mortality in cardiac surgery has been studied in a EuroSCORE multicenter study conducted in eight European countries [10,14]. The EuroSCORE is intended for external and general application. However, we are not aware of any study evaluating the performance of the EuroSCORE in other patient populations.

Our purpose was to first derive a valid risk prediction

* Corresponding author. Tel.: +358-17-173311; fax: + 358-17-173377.
E-mail address: otto.pitkanen@kuh.fi (O. Pitkänen).

1010-7940/00/$ - see front matter
PII: S1010-7940(00)00579-0

model for local use in our hospital. Secondly, since the performance of the recently published EuroSCORE [10,14] has not been tested in our patient material, we compared its performance and that of our newly derived model for predicting mortality and morbidity in our patient population.

2. Materials and methods

The study was conducted at Kuopio University Hospital and approved by the Ethics Committee. The need for informed consent was waived, because the data used for the study had already been collected for clinical purposes. The material consisted of 4592 patients having undergone cardiac surgery during a 5-year period between January 1, 1992, and December 31, 1996. In addition, data from 821 consecutive cardiac surgery patients operated between September 1, 1998 and May 31, 1999, were included. The patients operated without cardiopulmonary bypass were excluded from the study. Two of the investigators (S.R. and O.P.) retrospectively collected the clinical and physiological data. Data were entered into a computerized database. Morbidity was defined as one or more of the following 13 factors appearing postoperatively: haemodynamic problems indicating inotropic support or intra-aortic balloon pump, mechanical ventilator therapy required for longer than 24 h, serious gastrointestinal complication, anuria, stroke, multiorgan failure, resternotomy due to other cause than excessive bleeding, sepsis, pneumonia, mediastinitis, psychosis or remarkable confusion, a readmission to the intensive care unit (ICU) or a complicated clinical situation at discharge to another hospital. Perioperative mortality was defined as a death occurring within 30 days from the operation. The mortality data were obtained from the files of Statistics Finland, a central bureau registering all deaths of the Finnish citizens. Since we evaluated the accuracy of the EuroSCORE [10], we aimed to include the variables used in the EuroSCORE also into our models. The variables collected were the same as in the EuroSCORE with one exception. The preoperative pulmonary artery systolic pressure was missing in our retrospective database ($n = 4592$), but was included in the prospectively collected database ($n = 821$).

Predictive models were developed by logistic regression analyses. Firstly, the retrospective database of 4592 patients was utilized. Two thirds of the patients ($n = 3061$) were randomized to a derivation database, which was used to derive predictive models. The remaining one third ($n = 1531$; retrospective validation database) was used as an independent database for validation of the models. Secondly, the models were validated also in a prospectively collected database ($n = 821$; prospective validation database). Three outcome states were defined as follows: mortality, morbidity as described above, and length of stay (LOS) in the ICU for more than 2 days (ICU LOS > 2 days).

Three models to predict three outcome states were derived in the randomized retrospective derivation database ($n = 3061$) in two steps as follows. First, univariate analyses were performed in order to find out those predictive variables that were associated with the outcome states. Chi-square test or Fisher's exact test as appropriate were performed if the categorized predictive variables occurred at least in 0.5% of cases in the sample. Unpaired *t*-test was used to study for the differences of continuous variables. Predictive variables with a *P*-value of less than 0.20 in the univariate analyses were potentially eligible into logistic regression analysis. Second, the correlation between those variables that were significant in the univariate analyses was tested by Pearson's correlation (continuous variables) or Spearman's rank order correlation (categorized or nominal variables). Because selection of variables that correlate with each other may result in multicollinearity and overfitting of the model, only one clinically relevant variable was chosen in case of correlation ($P > 0.05$). A total of 17 variables describing the chronic health of the patient, 16 describing the preoperative status and examinations and 14 concerning the type and priority of the operation were screened in univariate analyses. The corresponding number of variables entered in the backward stepwise logistic regression analyses were 12, 16 and 13, respectively.

The significant variables ($P < 0.05$) after backward stepwise elimination formed the final predictive model including eight variables predicting mortality, 14 variables predicting morbidity, and 12 variables predicting the length of intensive care unit stay > 2 days.

All three models explaining mortality, morbidity or ICU LOS > 2 days were validated both in the retrospective ($n = 1531$) an in the prospective validation ($n = 821$) databases. Model calibration (precision) was evaluated by the Hosmer–Lemeshow goodness-of-fit statistics [15]. The discrimination abilities (accuracy) of the predictive models and the EuroSCORE were assessed with the area under the receiver-operating characteristic (ROC) curve [16].

The difference in the EuroSCORE between patients with and without mortality, morbidity and ICU LOS > 2 days was tested by the non-parametric Mann–Whitney *U*-test. Differences in patient characteristics, type and priority of operation, outcome variables and the predicted risks between the prospective and retrospective validation databases was performed by chi-square test, Fisher's exact test, unpaired *t*-test or Mann–Whitney *U*-test as appropriate. Statistical significance was defined as $P < 0.05$. The results are given as mean ± SD unless indicated otherwise. All the statistical procedures including randomization and ROC analyses were performed by SPSS 9.0 statistical package (SPSS Inc., Chicago, IL).

3. Results

Patient characteristics, type and priority of the operation and the observed outcome of the patients belonging either to the whole retrospective database ($n = 4592$) or the prospective validation database ($n = 821$) are presented in Table 1. When the prospective validation database ($n = 821$) and the retrospective validation database ($n = 1531$) were compared, the patients in the prospective database were older (63 ± 10 vs. 61 ± 9 years, $P < 0.0001$), the proportion of female patients was higher (30 vs. 26%, $P = 0.02$), the incidence of chronic renal failure was higher (3 vs. 1%, $P < 0.0001$), and ICU LOS was shorter (1.4 ± 1.9 vs. 1.9 ± 3.6 days, $P < 0.0001$). The proportion of patients undergoing only coronary artery bypass grafting (CABG) was similar (75.6 vs. 78.6%, not significant). The EuroSCORE, mortality or morbidity were not different in the two validation databases.

The predictive models and the independent contribution of each variable to the outcome states are presented in Table 2. There was a difference in the calculated risk of mortality and morbidity between the prospective and the retrospective validation databases (0.02 ± 0.05, median 0.01 vs. 0.02 ± 0.04, median 0.008; $P = 0.002$ and 0.17 ± 0.14, median 0.13 vs. 0.17 ± 0.15, median 0.15; $P < 0.0001$, respectively). The validation databases were not different in regard to the expected risk of ICU LOS > 2 days (0.08 ± 0.10, median 0.05 vs. 0.08 ± 0.10, median 0.04; not significant).

All the three predictive models calibrated well in the validation databases with the exception of the model predicting morbidity (Table 3). The discrimination abilities of our models and that of the EuroSCORE were similar (Table 4). The discriminative power of our model and the EuroSCORE in predicting morbidity was not very good since the areas under the ROC curves were below 0.75 both in the retrospective and prospective datasets (Table 4). In the prospective validation database ($n = 821$), the EuroSCORE was higher among non-survivors than survivors (6.7 ± 3.3 vs. 3.6 ±2.8, $P = 0.001$), among patients with morbidity than without morbidity (5.6 ± 2.9 vs. 3.1 ± 2.5, $P < 0.0001$), and among those having ICU LOS for more than > 2 days than among those with a shorter ICU LOS (6.7 ± 3.3 vs. 3.4 ± 2.6, $P < 0.0001$). The mortality and morbidity rates and the occurrence of ICU LOS > 2 days increased with the increasing EuroSCORE (Fig. 1). The rates of mortality, morbidity and ICU LOS > 2 days according to low, medium and high risk of the EuroSCORE in the validation databases are shown in Table 5.

Table 1
Clinical data, type of operation and observed outcome

Patient characteristics	Retrospective database ($n = 4592$)	Prospective database ($n = 821$)
Age (years)[a]	61 ± 9 (6–84)	63 ± 10 (21–84)
Gender (male/female) (%)	73/27	70/30
EuroSCORE[a]	3.2 ± 2.6 (0–15)	3.6 ± 2.8 (0–17)
Operation (%)		
CABG only	79.5	75.6
Combined CABG and valve procedure	9.5	9.6
Single valve procedure	6.3	10.0
Combined aortic and mitral valve procedure	0.5	0.4
Combined aortic and mitral valve procedure and CABG	0.6	0.4
Others	3.6	4.0
Surgical priority (%)		
Elective	70.3	65.0
Urgent	26.2	32.6
Emergent	3.5	2.4
Outcome		
30-day mortality (%)		
All	2.0	1.1
Coronary artery bypass graft only	1.2	0.9
Elective and urgent cases	1.5	1.0
Emergent cases	15.7	5.3
Morbidity (%)		
All	22.0	18.4
Elective and urgent cases	20.7	17.2
Emergent cases	60.4	68.4
Length of intensive care unit stay (days) (all)[a]	1.9 ± 3.9 (0–75)	1.4 ± 1.9 (0–29)
Length of postoperative hospital stay (days) (all)[a]	7.1 ± 5.2 (0–105)	6.3 ± 2.7 (0–35)

[a] Mean ± SD (range).

Table 2
Multivariate predictors of mortality, morbidity, and the length of stay in the intensive care unit[a]

Variable (unit)	Mortality				Morbidity				ICU LOS > 2 days			
	Coefficient	*P*-value	Odds ratio	95% CI	Coefficient	*P*-value	Odds ratio	95% CI	Coefficient	*P*-value	Odds ratio	95% CI
Age (years)	0.082	0.001	1.09	1.04–1.13	0.043	0.000	1.04	1.03–1.06	0.035	0.000	1.04	1.02–1.06
Female gender (0,1)					0.275	0.018	1.32	1.05–1.65	0.480	0.003	1.62	1.18–2.21
NYHA class (1–4)	0.313	0.000	1.37	1.17–1.59	0.252	0.002	1.29	1.10–1.51	0.397	0.001	1.49	1.18–1.87
Diabetes (0,1)	0.763	0.022	2.14	1.12–4.12					0.454	0.012	1.57	1.11–2.24
Previous stroke (0,1) (hemiparesis/hemiplegia or cerebellar infarct)					0.644	0.022	1.90	1.09–3.29	1.285	0.000	3.61	1.95–6.70
ASO in lower limbs (0,1) (occlusion of lower limb arteries)						n.s.			0.511	0.031	1.67	1.05–2.65
Number of previous AMIs (*n*)	0.428	0.005	1.54	1.14–2.07	0.280	0.000	1.32	1.16–1.51				
Previous inferior AMI (0,1)									0.610	0.001	1.84	1.27–2.66
Diuretic use (0,1)					0.298	0.023	1.35	1.04–1.74	0.415	0.016	1.51	1.08–2.12
Renal failure (0,1) (serum creatinine > 120 μmol/l preoperatively)					0.885	0.029	2.42	1.09–5.38				
LV ejection fraction (%)	−0.025	0.017	0.98	0.96–0.99	−0.016	0.000	0.99	0.98–0.99	−0.022	0.000	0.98	0.97–0.99
Pulmonary rales (0,1)					0.519	0.040	1.68	1.02–2.76				
CABG only (0,1)	−1.197	0.000	0.30	0.16–0.57	−0.649	0.001	0.52	0.36–0.75	−1.132	0.000	0.32	0.23–0.45
UAP and ongoing AMI (0,1)					2.092	0.013	8.09	1.56–42.1	1.401	0.037	4.06	1.09–15.1
Combined CABG and valve operation (0,1)					0.549	0.008	1.73	1.16–2.60				
Combined AVR and MVR (0,1)					1.520	0.009	4.57	1.47–14.2				
Combined AVR and MVR and CABG (0,1)	1.605	0.027	4.98	1.19–20.7								
Emergency operation (0,1)	1.465	0.001	4.33	1.78–10.5	0.734	0.016	2.08	1.14–3.80	0.959	0.001	2.61	1.35–5.03
Constant	−7.542				−3.957				−4.389			

[a] 0,1 refers to a dichotomous variable. ICU LOS, length of stay in the intensive care unit; CI, confidence interval; NYHA, New York Heart Association; ASO, obliterative arteriosclerosis; AMI, acute myocardial infarction; LV, left ventricular; CABG, coronary artery bypass grafting; UAP, unstable angina pectoris; AVR, aortic valve replacement; MVR, mitral valve replacement or reconstruction.

Table 3
Calibration (goodness-of-fit, Hosmer–Lemeshow) of the predictive models in retrospective ($n = 1531$) and prospective ($n = 821$) validation databases

	P-value
Mortality	
Retrospective validation data	0.466
Prospective validation data	0.320
Morbidity	
Retrospective validation data	0.002
Prospective validation data	0.082
Length of ICU stay > 2 days	
Retrospective validation data	0.401
Prospective validation data	0.484

4. Discussion

We derived three internal models to predict outcome and to allocate resources after cardiac surgery and compared their performance with the EuroSCORE. Our models predicting risk of mortality, morbidity and prolonged stay in the ICU fitted well in the validation databases, especially when assessed prospectively. In discriminating patients with a good outcome from those with poor outcome, our models and the EuroSCORE appeared to be equally accurate.

The unadjusted operative mortality among all our cardiac surgery patients (2.0 and 1.1% in the retrospective and prospective databases, respectively) and also among our patients with pure CABG (1.2 and 0.9% in the retrospective and prospective databases, respectively) was lower or equal as compared to previous reports [6,10,12,17,18]. The incidence of morbidity among our patients was 22 and 18% in the retrospective and prospective databases, respectively. Similar figures for morbidity have been published by Tuman et al., although the definitions for morbidity were somewhat different [5]. In our study 6.5 and 4.8% of CABG patients (in the retrospective and prospective databases, respectively) had prolonged ICU LOS (>2 days), whereas Wong et al. [9] reported that 17% of CABG patients stayed more than 48 h in the ICU after fast-track protocol.

There were differences in the patient material between the

Table 4
Discrimination ability of the present predictive models and that of the EuroSCORE system in retrospective ($n = 1531$) and prospective ($n = 821$) validation databases

	Area under receiver-operating characteristic curve (95% CI)	
	Predictive models	EuroSCORE
Mortality		
Retrospective validation data	0.82 (0.74–0.89)	0.81 (0.73–-0.89)
Prospective validation data	0.84 (0.70–0.99)	0.77 (0.63–-0.92)
Morbidity		
Retrospective validation data	0.73 (0.70–0.76)	0.70 (0.67–0.73)
Prospective validation data	0.74 (0.70–0.79)	0.74 (0.70–0.78)
Length of ICU stay of > 2 days		
Retrospective validation data	0.75 (0.71–0.80)	0.76 (0.71–0.80)
Prospective validation data	0.81 (0.76–0.87)	0.79 (0.73–0.84)

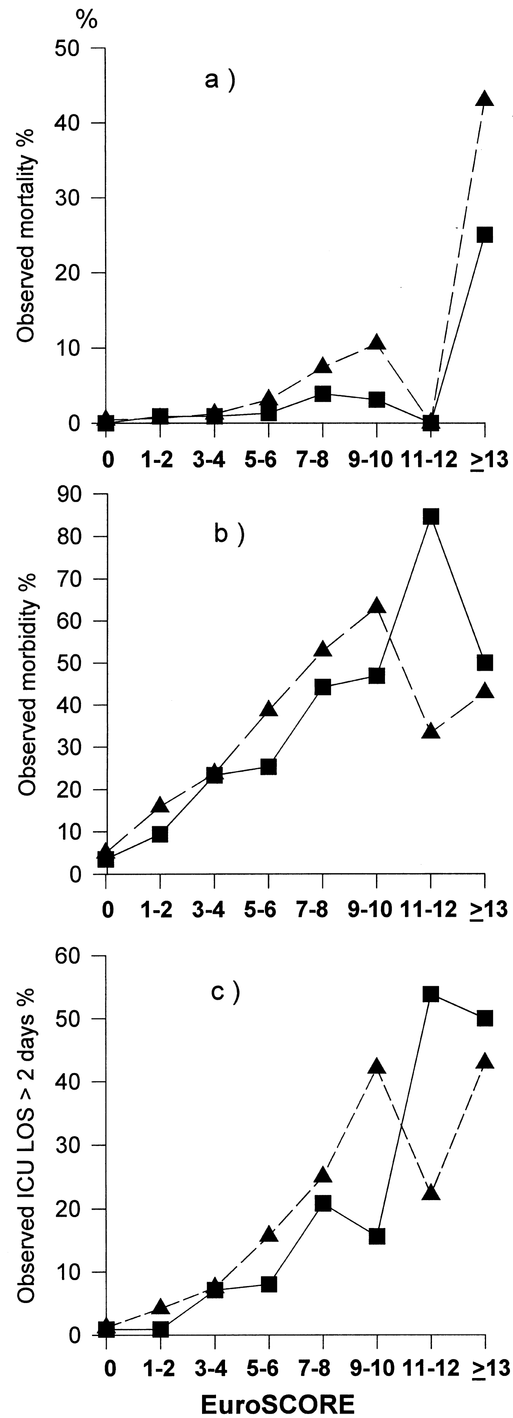

Fig. 1. Mortality (a), morbidity (b) and length of stay in the intensive care unit (ICU) > 2 days (c) in the prospective ($n = 821$) and the retrospective validation databases ($n = 1531$) vs. EuroSCORE. ■, prospective validation database; ▲, retrospective validation database.

Table 5
Mortality, morbidity and length of intensive care unit stay in the retrospective ($n = 1531$) and in the prospective ($n = 821$) validation database according to EuroSCORE category

	EuroSCORE 0–2 (low risk) *n* (%)	EuroSCORE 3–5 (medium risk) *n* (%)	EuroSCORE 6 or higher (high risk) *n* (%)
Retrospective validation database	714 (47)	548 (36)	269 (18)
Prospective validation database	326 (40)	309 (38)	86 (23)
Mortality (%)			
Retrospective validation database	0.6	1.3	7.8
Prospective validation database	0.3	0.6	3.2
Morbidity (%)			
Retrospective validation database	12.2	27.2	48.7
Prospective validation database	7.4	23.3	42.5
ICU LOS > 2 days (%)			
Retrospective validation database	3.2	9.1	25.3
Prospective validation database	0.9	6.5	19.9

periods of the two databases, i.e. from 1992–1996 to 1998–1999. The patients in the more recent database were older, the proportion of female patients was higher and there were less chronic renal failure and pure CABG operations. Also the predicted risk for mortality was higher. However, in our more recent, i.e. prospective database the ICU LOS was shorter. This may reflect a change in our routine postoperative care towards a fast-track practice or better response to given case. A similar trend toward an increasing proportion of high-risk patients with simultaneous decrease in ICU LOS has been previously reported from North America [8,12,19].

The multivariate predictors of outcome, such as age, gender, chronic co-morbidities, left ventricular ejection fraction, priority and type of operation in our model were quite the same as previously published [1–3,8,17]. The predictors we utilized can be regarded as objective and reproducible indicators of outcome with the exception of New York Heart Association (NYHA) class which varies somewhat according to the assessor. Also diuretic therapy as a risk factor may be dependent on physician and institutional preferences. ICU LOS reflects the overall use of resources and has been predicted also in other studies [4,6,9]. The time limit chosen for our study was based on the median ICU LOS in the derivation database and is the same as in a previous report [9].

The performance, i.e. discrimination and calibration, of our models in predicting mortality, morbidity or prolonged ICU LOS was comparable to other risk indexes that have been created for cardiac surgery patients [3,6–8,10]. In the study by Wong et al. [9] the discrimination abilities of models predicting ICU stay lasting two days or longer were somewhat better (area under the ROC curve 0.85) than those in our model (0.75–0.81). Our models were all based on preoperative factors, while the models by Wong et al. [9] included also postoperative factors such as inotrope use or use of intra-aortic balloon pump thereby increasing the predictive ability of their models. In spite of the differences in patient characteristics and outcome between the two periods of 1992–1996 and 1998–1999, the models derived from 1992–1996 data were accurate enough in the more recent validation database from 1998–1999. However, predictive models should be episodically updated and recalibrated to ensure their optimal performance because there tends to be evolution in medical and surgical techniques along time [8,11].

Both our model and the EuroSCORE were not very accurate in predicting morbidity. This may be due to the numerous conditions defined as factors for morbidity. The models might have performed better in the prediction of one isolated event rather than 13 events.

Our models are equations that provide an estimated risk of death, morbidity or prolonged ICU LOS for each patient. The EuroSCORE is a simple additive score designed to provide the physician and the patient a simple tool to estimate the risk of death. It is an estimated percentage probability of death and very close to the operative mortality observed in the pilot program.

The EuroSCORE was divided into three risk groups [10]. The thresholds for these risk groups were based on the equal size of those three groups. In our material the proportion of patients with low EuroSCORE was higher (40–47%), and correspondingly, the proportion of patients with high EuroSCORE was lower (18–23%) than in the original validation database of the EuroSCORE. This finding suggests that patients in our study had, on the average, lower EuroSCOREs than the patients in the original EuroSCORE data. Further, our patients were less severely ill having lower risk of death than the European cardiac surgical patients in general. Nevertheless, the observed mortality rates in our validation databases were lower than in the EuroSCORE study in each of the three risk categories (0.3–0.6 vs. 0.8%, 0.6–1.3 vs. 3% and 3.2–7.8 vs. 11.2% in low, medium and high risk categories, respectively). Thus, it may be argued that the outcome and presumably the quality of care provided in our institution might have been better among our patient material than among that of the EuroSCORE study. Whether or not this was due to some differences in the operation-related factors between our study and the EuroSCORE study, remains uncertain. In the EuroSCORE study, there were less patients undergoing pure CABG operation, but on the other hand, more elective operations and slightly less urgent operations than in our study.

In spite of the differences in the patient characteristics between our databases and that of the EuroSCORE, the latter performed quite well in our databases. The discrimination ability of the EuroSCORE was comparable with that of our models. In our prospective validation database, 25% of patients with a EuroSCORE of 13 or higher died (Fig. 1). The association between mortality and EuroSCORE values was much less clear in patients with EuroSCORE values 11–12 or lower. In contrast, the incidence of morbidity and prolonged ICU LOS seemed to increase with the increasing score. However, in the highest categories the correlation was less evident. This could be explained by the fact that these categories included only small numbers of patients.

An advantage of our model is the low number of variables needed in the risk calculation. Our models contained variables from eight to 14, while the number of variables required for calculation of the EuroSCORE is 17. The number of variables in our models seems appropriate because including too many variables in risk indexes not only increases cost and errors but also may result in statistical overfitting and instability [8]. A part of the variables included in our models and in the EuroSCORE (age, gender, previous stroke, arteriosclerosis in lower limbs, renal failure, left ventricle ejection fraction, unstable angina, type and priority of operation) were the same, but some of them were differently defined as predictive factors in the regression equations. For instance, age and left ventricle ejection fraction were handled as continuous variable in our study (Table 2) while in the EuroSCORE study age was divided into five years intervals and ejection fraction into two classes. In addition, our definition for neurologic dysfunction was not as exact as in the EuroSCORE study (our study: previous hemiplegia/hemiparesis or cerebellar infarct and EuroSCORE study: disease severely affecting ambulation or day-to-day functioning), and correspondingly, the definitions for ASO (our study: anamnestic or angiographic occlusion of lower limb arteries and EuroSCORE study: claudication, carotid occlusion or $> 50\%$ stenosis, previous or planned intervention on the abdominal aorta, limb arteries or carotids) and renal failure (our study serum creatinine > 120 μmol/l and the EuroSCORE study: serum creatinine > 200 μmol/l preoperatively) were also different. Therefore, the weights (odds ratios) for the variables in our study are not comparable with the EuroSCORE study. In addition, the different weights may be due to different frequencies of these variables in the derivation databases and the differences in patient material and outcome [10,14]. These factors may also explain the slightly unstable relationship between the observed outcome states in our material and the EuroSCORE and the skewed distribution of patients into three EuroSCORE risk categories in our databases. Due to these differences the comparison of our local models and the EuroSCORE should be undertaken with caution.

The local models and the EuroSCORE may not be mutually exclusive, rather they could be collected and used to supplement each other. However, before any predictive index is used in intra-institutional outcome prediction or quality control, its performance should be carefully evaluated in each institution. In our institution the EuroSCORE performed quite accurately. Therefore our finding does not support derivation of a local model for every cardiac surgery center. In addition to quite correct prediction of risk of death, our local model provided also a well calibrated estimation of probability of prolonged ICU stay. Thus it may have additive value especially for our local resource allocation purposes.

If the discrimination and calibration of a predictive index have proved to be appropriate, it can be used in selection of patients for operative treatment in borderline cases and to help physicians to prepare for oncoming complications [18]. Accurate prediction of ICU LOS or length of hospital stay may guide in allocating resources. Morbidity, mortality, and the ICU LOS can be predicted quite reliably on the grounds of preoperative information and the type of operation. Whether the models would be more precise, if intraoperative predictors were added [20], has to be tested in future studies. If the information needed in such predictive models could be stored into database as a part of the normal documentation during the treatment process by using patient data management systems, the index would be easily obtained and more applicable in clinical situations. However, it is to be stressed that clinicians have to be very cautious when applying any predictive indexes to individual patients [17].

In conclusion, while our models seem to provide a fine-tuned method in resource allocation and quality assurance purposes in local use in our hospital, the EuroSCORE may be appropriate in categorizing patients undergoing cardiac surgery in clinical trials and in interinstitutional comparisons.

Acknowledgements

This study has been supported by institutional grant from Kuopio University Hospital.

References

[1] Parsonnet V, Dean D, Bernstein AD. A method of uniform stratification of risk for evaluating the results of surgery in acquired adult heart surgery. Circulation 1989;79(1):3–12.

[2] Higgins TL, Estafanous FG, Loop FD, Blum JM, Paranandi L. Stratification of morbidity and mortality outcome by preoperative risk factors in coronary artery bypass patients. A clinical severity score. J Am Med Assoc 1992;267(17):2344–2348.

[3] O'Connor GT, Plume SK, Olmstead EM, Coffin LH, Morton JR, Maloney CT, Nowicki ER, Levy EG, Tryzelaar JF, Hernandez F, Adrian L, Casey KJ, Bundy D, Soule DN, Marrin CAS, Nugent

WC, Charlesworth DC, Clough R, Katz S, Leavitt BJ, Wenneberg JE. Multivariate prediction of in-hospital mortality associated with coronary artery bypass graft surgery. Circulation 1992;85:2110–2118.

[4] Higgins TL, Starr NJ, Estafanous FG. Predicting prolonged intensive care unit length of stay following coronary artery bypass. Anesth Analg 1996;82:S176.

[5] 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;102:23–44.

[6] Tu JV, Jaglal SB, Naylor CD. Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery. Circulation 1995;91:677–681.

[7] Kurki T, Kataja M. Preoperative prediction of postoperative morbidity in coronary artery bypass grafting. Ann Thorac Surg 1996;61:40–45.

[8] Ivanov J, Tu JV, Naylor D. 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.

[9] 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. Anesthesiology 1999;91:936–944.

[10] Nashef SA, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation. Eur J Cardiothorac Surg 1999;16:9–13.

[11] Krumholz HM. Mathematical models and the assessment of performance in cardiology. Circulation 1999;99:2067–2069.

[12] Edwards FH, Clark RE, Schwartz M. Coronary artery bypass grafting: the Society of Thoracic Surgeons national database experience. Ann Thorac Surg 1994;57:12–19.

[13] Hattler BG, Armitage JM, Hardesty RL. Risk stratification using the Society of Thoracic Surgeons program. Ann Thorac Surg 1994;58:1348–1352.

[14] Roqués 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.

[15] Hosmer DW, Lemeshow S. Applied logistic regression. New York: Wiley, 1989.

[16] Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839–843.

[17] Higgins TL. Quantifying risk and assessing outcome in cardiac surgery. J Cardiothorac Vasc Anesth 1998;12:330–340.

[18] Martinez-Alario J, Tuesta ID, Plasencia E, Santana M, Mora ML. Mortality prediction in cardiac surgery patients. comparative performance of parsonnet and general severity systems. Circulation 1999;99:2378–2382.

[19] Gover FL. The Society of Thoracic Surgeons national database: status and future directions. Ann Thorac Surg 1999;68:367–373.

[20] Reich DL, Bodian CA, Krol M, Kuroda M, Osinski T, Thys DM. Intraoperative hemodynamic predictors of mortality, stroke, and myocardial infarction after coronary artery bypass surgery. Anesth Analg 1999;89:814–822.