ELSEVIER

European Journal of Cardio-thoracic Surgery 17 (2000) 400–406

EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY

www.elsevier.com/locate/ejcts

Risk stratification in heart surgery: comparison of six score systems☆

Hans J. Geissler*, Philipp Hölzl, Sascha Marohl, Ferdinand Kuhn-Régnier, Uwe Mehlhorn, Michael Südkamp, E. Rainer de Vivie

Department of Cardiothoracic Surgery, University of Cologne, Joseph-Stelzmann-Strasse 9, 50924 Cologne, Germany

Received 6 September 1999; received in revised form 25 January 2000; accepted 8 February 2000

Abstract

Objective: Risk scores have become an important tool in patient assessment, as age, severity of heart disease, and comorbidity in patients undergoing heart surgery have considerably increased. Various risk scores have been developed to predict mortality after heart surgery. However, there are significant differences between scores with regard to score design and the initial patient population on which score development was based. It was the purpose of our study to compare six commonly used risk scores with regard to their validity in our patient population. **Methods**: Between September 1, 1998 and February 28, 1999, all adult patients undergoing heart surgery with cardiopulmonary bypass in our institution were preoperatively scored using the initial Parsonnet, Cleveland Clinic, French, Euro, Pons, and Ontario Province Risk (OPR) scores. Postoperatively, we registered 30-day mortality, use of mechanical assist devices, renal failure requiring hemodialysis or hemofiltration, stroke, myocardial infarction, and duration of ventilation and intensive care stay. Score validity was assessed by calculating the area under the ROC curve. Odds ratios were calculated to investigate the predictive relevance of risk factors. **Results**: Follow-up was able to be completed in 504 prospectively scored patients. Receiver operating characteristics (ROC) curve analysis for mortality showed the best predictive value for the Euro score. Predictive values for morbidity were considerably lower than predictive values for mortality in all of the investigated score systems. For most risk factors, odds ratios for mortality were substantially different from ratios for morbidity. **Conclusions**: Among the investigated scores, the Euro score yielded the highest predictive value in our patient population. For most risk factors, predictive values for morbidity were substantially different from predictive values for mortality. Therefore, development of specific morbidity risk scores may improve prediction of outcome and hospital cost. Due to the heterogeneity of morbidity events, future score systems may have to generate separate predictions for mortality and major morbidity events.

Keywords: Cardiac surgery; Mortality; Morbidity; Risk score; Risk factor

1. Introduction

Preoperative risk scores are an essential tool for risk assessment, cost–benefit analysis, and the study of therapy trends. Various score systems have been developed to predict mortality after adult heart surgery [1–9]. Although all of these score systems are based on patient derived data, such as age, gender, comorbidity, and so forth, there are considerable differences between scores with regard to their design and validity. As quality control and cost-benefit analysis have gained new relevance with recent developments in the health care system, selection of appropriate score systems for the evaluation of hospital performance has become an important issue. It was the purpose of our study to compare six commonly used preoperative risk scores for heart surgery with regard to their predictive values and clinical applicability for our patient population. Although most of the selected score systems were primarily designed to predict mortality, postoperative morbidity has been acknowledged as the major determinant of hospital cost and quality of life after surgery [10]. Therefore, we analyzed the selected risk scores not only with regard to their predictive value for mortality, but for postoperative morbidity as well.

2. Patients and methods

All adult patients undergoing heart surgery with cardiopulmonary bypass at the University of Cologne between September 1, 1998 and February 28, 1999 were prospectively scored according to the initial Parsonnet [1], Cleve-

☆ Presented at the 13th Meeting of the European Association for Cardiothoracic Surgery, Glasgow, Scotland, UK, September 5–8, 1999.

* Corresponding author. Tel.: +49-221-478-4128; fax: +49-221-478-4186.

E-mail address: hans.geissler@medizin.uni-koeln.de (H.J. Geissler)

1010-7940/00/$ - see front matter
PII: S1010-7940(00)00385-7

land Clinic [2], French [4], Euro [7,8], Pons [6], and Ontario Province Risk (OPR) score [5]. Scores were selected with regard to their acceptance in the literature and clinical applicability. Scoring was performed by assigned authors (P.H. and S.M.). Heart transplant recipients and patients operated on beating heart without cardiopulmonary bypass (off-pump surgery) were excluded from the study.

Table 1 summarizes the score items, which were evaluated by the six score systems.

Follow-up was continued for 30 days postoperatively.

Table 1
Risk score items

	Initial Parsonnet score	Cleveland Clinic score	French score	Euro score	Pons score	Ontario Province Risk score
Patient data						
Age	✓	✓	✓	✓	✓	✓
Gender	✓			✓		✓
Body weight	✓	✓				
Cardiac						
Unstable agina				✓		
Aortic stenosis	✓	✓				
Active endocarditis				✓		
Congenital heart defect	✓					
Hypertension, arterial	✓					
Hypertension, pulmonary	✓			✓		
LV aneurysm	✓				✓	
LV ejection fraction	✓	✓	✓	✓		✓
Mitral insufficiency		✓				
Myocardial infarction (MI)			✓	✓	✓	
NYHA					✓	
Post MI VSD			✓	✓		
Ventricular tachycardia/ fibrillation			✓	✓		
Pulmonary						
Asthma	✓			✓		
COPD		✓		✓		
Renal						
Dialysis	✓		✓			
Creatinine		✓	✓	✓	✓	
Acute renal failure	✓			✓		
Other						
Anemia		✓				
Diabetes	✓	✓				
Liver disease					✓	
History of TIA, stroke		✓		✓		
Paraplegia	✓					
Pacemaker	✓					
Vascular						
Aortic dissection, acute			✓			
Peripheral arterial disease				✓		
History of vascular surgery		✓		✓		
Preoperative						
Ventilation			✓	✓	✓	
IABP	✓			✓		
Inotropes				✓		
Resuscitation				✓		
Cardiogenic shock	✓				✓	
Operation						
Combined surgery	✓		✓	✓	✓	✓
Urgent/emergency	✓	✓		✓	✓	✓
Reoperation	✓	✓	✓	✓	✓	✓

The majority of patients was discharged or transferred to another hospital before the end of the 30-day period. In these patients follow-up was established by a questionnaire which was mailed to the patient. The patient's general practitioner was contacted in case of missing information.

The following points of outcome were investigated:

- death within 30 days of surgery;
- need for a mechanical assist device (intra-aortic balloon pump, ventricular assist device, or extracorporeal membrane oxygenation), implanted during or after surgery;
- postoperative renal failure requiring hemodialysis or hemofiltration;
- stroke (verified by neurological consult or CT-scan);
- intra- or postoperative myocardial infarction, verified by positive CK-MB in combination with either typical ECG changes or echocardiographically detected new left ventricular wall motion abnormalities;
- any return to surgery (rethoracotomy for bleeding, bypass revision, wound revision);
- prolonged ventilation (>48 h);
- prolonged stay in the intensive care unit (>6 days).

For the purpose of the study we defined morbidity by the above-mentioned points of outcome with the exception of death within 30 days. Morbidity was analyzed for the overall patient population.

2.1. Statistical analysis

Data are presented as absolute numbers, mean ± standard deviation, or percentages. Data acquisition of the more than 40 000 data entries was performed using Microsoft Access and Excel, version 97. Data analysis was performed using the SPSS software package, version 8.01. Nominal data were analyzed using χ^2 or, where appropriate, Fisher's exact test. Receiver operating characteristics (ROC) curves were plotted for the different score systems and the area under the ROC curve was calculated as an index for the predictive value of the model (Fig. 1). Areas under ROC curves were compared according to the statistical approach suggested by Hanley and McNeil [11] using the MEDCALC 5.0 software package. A Bonferroni-correction was used to correct for multiple comparisons. To analyze the predictive value of specific risk factors or score items we calculated the according odds ratios. A P-value of less than 0.05 was considered significant.

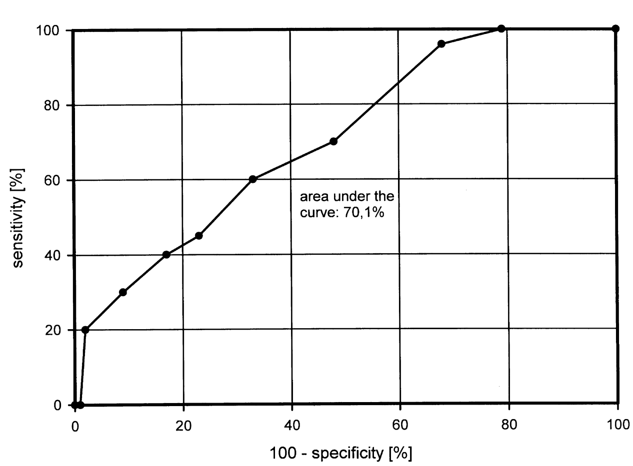

Fig. 1. Example for receiver operating characteristics (ROC) curve: mortality curve for Ontario Province Risk score.

3. Results

Five hundred and five patients were prospectively scored and operated on during the study period. Follow-up was completed in 504 patients (99.8%). Mean age was 64 ± 10.5 years. 25.6% were female.

Table 2 shows the distribution of surgeries performed. Fig. 2 shows the distribution of risk factors among the study patients.

The actual 30-day mortality was 4%. The Cleveland, French and OPR scores predicted mortality between 3.5 and 4.9%, whereas mortality was considerably overestimated by the predictions of the Parsonnet, Euro, and Pons scores (Fig. 3).

Morbidity consisted of 180 events, as defined in Section 2, which occurred in 90 patients (17.9%). Fig. 4 shows the distribution of postoperative morbidity.

ROC curves were plotted separately for mortality and morbidity for each score. The greatest area under the curve (i.e. the highest predictive value) showed the mortality ROC curve of the Euro score with 78.6%. However, differences between areas under the curve were statistically not significant and all scores showed areas under the curve greater than 70%. Areas under the curve for morbidity were considerably lower than for mortality for all scores (Table 3).

Calculation of odds ratios showed that the predictive values for well-accepted risk factors such as diabetes, hypertension, obesity, unstable angina, and female gender were not significant. However, the predictive values of peripheral arterial insufficiency, decreased left ventricular function, a history of vascular surgery, older age, and preo-

Table 2
Distribution of cardiac surgical procedures

	No. of patients	Percentage
CABG	349	69.2
AVR	60	11.9
MVR	18	3.6
CABG + AVR	45	8.9
CABG + MVR	9	1.8
CABG + double valve	3	0.6
CABG + triple valve	1	0.2
CABG + aortic aneurysm	2	0.4
Double valve	7	1.4
Aneurysm ascending aorta	10	2
Σ	504	100

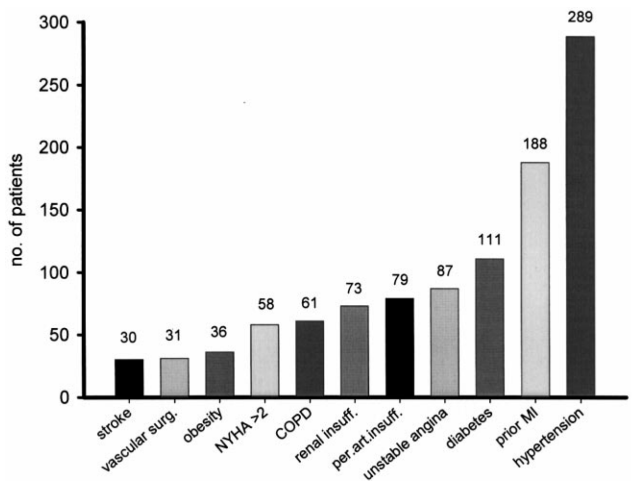

Fig. 2. Distribution of risk factors in the study population. MI, myocardial infarction.

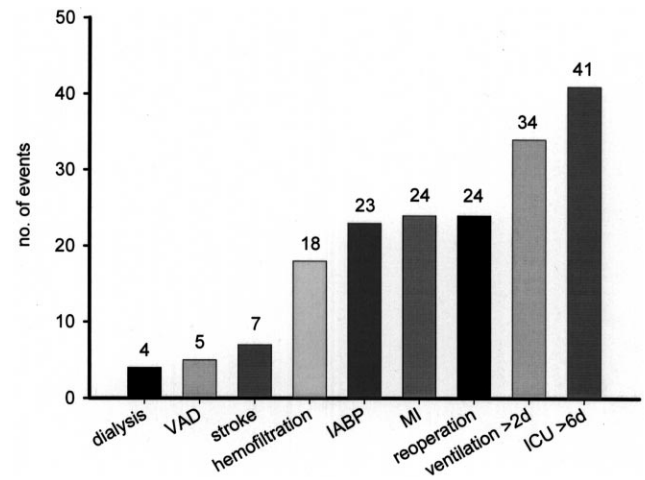

Fig. 4. Distribution of postoperative morbidity. VAD, ventricular assist device; IABP, intra-aortic balloon pump; MI, myocardial infarction; ICU, intensive care unit.

peratively increased serum creatinine were statistically significant with regard to mortality (Table 4).

4. Discussion

Analysis of patient outcome has gained increasing importance, as institutions, health care providers, and patients demand statistically sound data on risk and prognosis for specific procedures and therapies [12]. In particular, cost-intensive surgical procedures such as coronary artery bypass graft (CABG) surgery have received great attention with regard to cost–benefit analysis and comparison of mortality rates among institutions. As patient populations may differ significantly between institutions and countries, it became obvious that comparison of absolute numbers, such as mortality rates, was not feasible [1,10]. Various risk scores have been developed to correct for differences in patient population and to allow comparison of actual outcome to predicted outcome [1–9]. However, there are significant differences between scores with regard to the initial patient population on which score design was based. The clinical data base used for score development may have been derived from a single [1,2] or a number of institutions [4–7], from one country [1–6] or a number of neighboring countries [7]. Further differences include retrospective versus prospective data collection and whether a prospective validation study was completed following the score design (Table 5). Besides the Euro score, none of the selected scores was developed under inclusion of heart centers in Germany. Thus, one important goal of our study was to compare the selected risk scores with regard to their validity for our patient population. The study was supposed to supply data which may assist in selecting the most appropriate score for our institution.

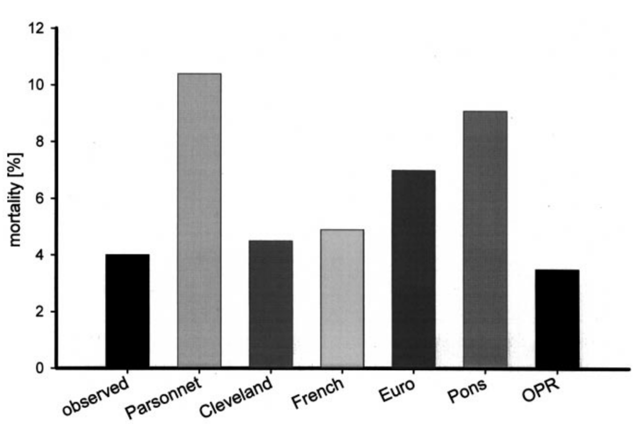

Fig. 3. Observed mortality in comparison to score predicted mortality. OPR, Ontario Province Risk score.

Controversial among investigators is the most appropriate statistical model for score development. Among the applied statistical tools are the calculation of simple odds ratios, logistic regression analysis [1–9], and Bayesian models [13]. Logistic regression analysis has been applied most frequently and results of various investigators show that risk scores with good predictive value can be developed using this statistical model.

Analysis of ROC curves yielded results for areas under the curve which are in fairly good agreement with those

Table 3
Validity of scores (areas under ROC)

	Area under ROC (%)	
	Mortality	Morbidity
Initial Parsonnet score	75.5	64.4
Cleveland Clinic score	73.1	68.6
French score	71.9	63.6
Euro score	78.6	63.8
Pons score	74.5	67.6
Ontario Province Risk score	70.1	62.1

Table 4
Predictive value of score items[a]

	Mortality			Morbidity		
	Odds ratio	*P*	95% confidence interval	Odds ratio	*P*	95% confidence interval
Vascular surgery	7.9	0.001	2.8–22.2	1.9	0.14	0.8–4.3
Peripheral vascular insufficiency	4.8	0.001	1.9–12.1	1.5	0.17	0.8–4.3
EF < 55%	4.8	0.001	1.9–11.8	1.5	0.15	0.9–2.7
Creatinine > 1.6 mg/dl	5.8	0.012	1.8–19	1.6	0.5	1.7–9.5
Age > 65 years	3.7	0.017	1.2–10.8	1.2	0.51	0.7–1.9
COPD	3.6	0.019	1.3–9.7	1.3	0.004	1.3–4.5
Combined surgery	3.2	0.022	1.2–8.2	2.3	0.004	1.3–4
TIA or stroke	4.4	0.025	1.4–14.1	2	0.13	0.8–4.6
NYHA IV	3.6	0.028	1.2–16	4.9	0.0001	2.6–9.3
Emergency	4.3	0.051	1.2–16	13.3	0.0001	5.2–33.6
Reoperation	3.2	0.06	1.0–10.1	2.2	0.04	1.0–4.8
V-tach, V-fib	5.7	0.066	1.1–29.1	1.2	0.7	0.2–5.4
LV-aneurysm	4	0.12	0.8–19.1	1.92	0.28	0.6–6.2
BMI > 33	3.5	0.14	0.7–16.3	1.6	0.5	0.5–3.1
Diabetes	3	0.23	1.2–7.6	2	0.009	1.1–3.3
MI < 30 days	2.3	0.24	0.7–8.3	2.3	0.058	0.9–5.3
Hypertension	1.8	0.24	0.7–4.7	1.1	0.63	0.7–1.8
Female gender	1.6	0.34	0.6–4.1	1.4	0.181	0.8–2.4
EF < 30%	2.5	0.36	0.3–20.5	4.5	0.021	1.3–15.1
Unstable angina	0.8	1	0.2–2.9	1	0.96	0.5–1.8

[a] EF, ejection fraction; V-tach, -fib, ventricular tachycardia, fibrillation; BMI, body mass index; MI, myocardial infarction.

reported in the literature [4,7,14,15]. With regard to mortality, the highest predictive value was calculated for the Euro score (Table 3). Among the selected scores, the Euro score has been the one most recently developed and involved the highest number of patients and institutions for its development, collecting data from 132 centers in eight European countries. Although differences between scores for areas under the ROC curve were statistically not significant, it is important to note that the selected score systems in this study give no information on the minimally required sample size for accurate predictions. Therefore, statistical comparisons based on larger patient numbers might come to different results. With regard to mortality, all of the selected scores showed areas under the curve greater than 70% and qualified therefore as applicable models, as an area under the curve greater than 70% is usually considered to be associated with a good predictive value [16].

Although in our study the area under the curve for the initial Parsonnet score was 75.5%, indicating a good correlation between increasing score value and mortality, overall mortality was considerably overestimated by this score. The data base for the initial Parsonnet score is now older than 12 years, and it seems likely that its predictive value was lessened by advances in surgical and medical therapy achieved during this period of time. As this process would apply to any score system over time, revalidation of score

Table 5
Design of selected risk scores

	Developed to predict	Center	Score development	Score validation	Statistics
Initial Parsonnet score 1989	Mortality	Single center/ USA	Retrospective: 3500 patients	Prospective: 1332 patients	Univariate and logistic regression analysis
Cleveland Clinic score 1992	Mortality/ morbidity	Single center/ USA	Retrospective: 5051 patients	Prospective: 4069 patients	Univariate and logistic regression analysis
French score 1995	Mortality/ morbidity	Multicenter/ France	Prospective: 7181 patients	–	Univariate and logistic regression analysis
Euro score 1999	Mortality	Multicenter/ Europe	Prospective: 19 030 patients	–	Univariate and logistic regression analysis
Pons score 1996	Mortality	Multicenter/ Spain	Prospective: 916 patients	Prospective: 392 patients	Univariate and logistic regression analysis
Ontario Province Risk score 1995	Mortality/ morbidity	Multicenter/ Canada	Retrospective: 6213 patients	Retrospective: 6885 patients	Univariate and logistic regression analysis

items at regular intervals seems warranted. However, in the case of the Parsonnet score we did not apply the modified Parsonnet score [3], because the clinical applicability of this complex score with several rather subjective items appears to be limited [17].

Mortality has been referred to as the most important performance indicator in heart surgery [18] and is the most frequently reported outcome parameter in evaluating risk scores. A clear advantage of assessing mortality is that it leaves little room for subjectivity in data acquisition, whereas objective parameters for morbidity are harder to define. Because morbidity is comprised of parameters as heterogeneous as need for a mechanical assist device or reoperation for bleeding, it appears to be difficult to find common risk factors for the prediction of these events. Furthermore, the impact of specific postoperative events, such as ventricular arrhythmia or prolonged ventilation, on long-term outcome remains controversial [10]. However, for postoperative events such as stroke, the impact on health care cost and quality of life has been widely acknowledged. Therefore, risk stratification for at least certain morbidity events appears to be desirable.

Our data show for all selected scores a substantially lower predictive value for morbidity than for mortality. The highest predictive value for morbidity shows the Cleveland Clinic score. However, when comparing these results one has to consider that morbidity parameters selected by us were different from those originally used by score developers. In addition, the Parsonnet, Euro and Pons scores were not designed for prediction of morbidity. Furthermore, analysis of odds ratios show that for most risk factors the predictive value for mortality differs considerably from that for morbidity (Table 4). Thus, we conclude that the statistical weight of certain risk factors may be different for the prediction of morbidity than for prediction of mortality. As morbidity is comprised of heterogeneous events, even a single risk factor may have significantly different odds ratios for various morbidity events.

Analyzing six different score systems for our patient population, the Euro score yielded the best predictive value for mortality. Predictive values for morbidity were substantially lower in all score systems, even in those specifically designed for the prediction of morbidity. Development of specific morbidity scores appears to be desirable for prediction of hospital cost and quality of life after surgery. However, due to the heterogeneity of morbidity events, a statistically sound prediction of overall morbidity is difficult to achieve. Future score systems may generate separate predictions for mortality and major morbidity events by adjusting for the different odds ratios of risk factors calculated with regard to mortality and various morbidity events.

Acknowledgements

The authors wish to thank Dr Wassmer, Institute for Medical Statistics and Documentation at the University of Cologne, for his statistical advice.

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 disease. Circulation 1989;79(suppl I):I3–I12.

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

[3] Hattler BG, Madia C, Johnson C, Armitage JM, Hardesty RL, Kormos RL, Payne DN, Griffith BP. Risk stratification using the Society of Thoracic Surgeons program. Ann Thorac Surg 1994;52:1348–1352.

[4] Roques F, Gabrielle F, Michel P, de Vincentiis C, David M, Baudet E. Quality of care in adult heart surgery: proposal for a self-assessment approach based on a French multicenter study. Eur J Cardio-thorac Surg 1995;9:433–440.

[5] 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;91:677–684.

[6] Pons JMV, Granados A, Espinas JA, Borras JM, Martin I, Moreno V. Assessing open heart surgery mortality in Catalonia (Spain) through a predictive risk model. Eur J Cardio-thorac Surg 1997;11:415–423.

[7] 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 0 patients. Eur J Cardio-thorac Surg 1999;15:816–823.

[8] Nashef SAM, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardio-thorac Surg 1999;16:9–13.

[9] Tremblay NA, Hardy JF, Perault J, Carrier M. A simple classification of the risk in cardiac surgery: the first decade. Can J Anaesth 1993;40:103–111.

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

[11] 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.

[12] Iezzoni LI. The risks of risk adjustment. J Am Med Assoc 1997;278:1600–1607.

[13] L'Italien PSD, Hendel RC, Leppo JA, Cohen MC, Fleisher LA, Brown KA, Zarich SW, Cambria RP, Cutler BS, Eagle KA. Development and validation of a Bayesian model for perioperative cardiac risk assessment in a cohort of 1081 vascular surgical candidates. J Am Coll Cardiol 1996;27:779–786.

[14] Orr RK, Maini BS, Sottile FD, Dumas EM, O'Mara P. A comparison of four severity-adjusted models to predict mortality after coronary artery bypass graft surgery. Arch Surg 1995;130:301–306.

[15] Pons JMV, Espinas JA, Borras JM, Moreno V, Martin I, Granados A. Cardiac surgical mortality: comparison among different additive risk-scoring models in a multicenter sample. Arch Surg 1998;133:1053–1057.

[16] Swets JA. Measuring the accuracy of diagnostic systems. Science 1988;240:1285–1293.

[17] Gabrielle F, Roques F, Michel P, Bernard A, de Vincentiis C, Roques X, Brenot R, Baudet E, David M. Is the Parsonnet's score a good predictive score of mortality in adult cardiac surgery: assessment by a French multicentre study. Eur J Cardio-thorac Surg 1997;11:406–414.

[18] Nashef SAM, Carey F, Silcock MM, Oommen PK, Levy RD, Jones MT. Risk stratification for open heart surgery: trial of the Parsonnet system in a British hospital. Br Med J 1992;305:1066–1067.

Appendix A. Conference discussion

Dr B. Messmer (*Aachen, Germany*): Would you make a patient selection and would you deny a patient the operation because your score shows a high risk?

Dr Geissler: I don't think that risk scores are suitable to make decisions on individual patients.

Dr Messmer: That is what I wanted to hear and I think that is to be underlined.

Dr Geissler: I think they are a very good tool to detect changes in patient populations, to study trends in therapy, but I don't think they are a good tool to make decisions on individual patients.

Dr F. Grover (*Denver, CO, USA*): One question I have is whether you update your indices or risk coefficients every year or two? We found in our own STS database that we have to update our risk coefficient frequently, because there are changes over time in how the different risk factors are weighted, at least according to mortality.

Dr Geissler: I think this is a very important factor. We have seen in this study that the initial Parsonnet score overestimated mortality vastly; however, the initial Parsonnet score was the oldest of the scores applied, it was designed in 1989, but if you look at the ROC curve analysis, the ROC curve analysis is pretty decent for the Parsonnet score despite the pretty poor prediction of mortality. So I think the reason for this is probably that the score is 10 years old and we applied the initial version.

Dr P. Sergeant (*Leuven, Belgium*): I greatly appreciated the effort, and I think one of the last comments made by the author is very important. We should realize that the prediction for every event requires a different scoring system. An additional comment is that we are not scoring the quality of care, we are only scoring the risk of care, and forget completely the late benefit of care. So, related to Dr Messmer's comments, we should definitely not decide about an indication for surgery based on the scoring, because we are not having any insight into the benefit of surgery.

One observes more and more, in abstracts and in publications, mortality prediction systems evaluated for their accuracy in predicting morbidity. We should absolutely split them up if one wants to get good insights. An acceptable ROC is no final proof of its applicability. Different events are often defined by the same incremental risk factors but the coefficients and the transformations of the variables will differ from event to event. One scoring system will never adequately define every event.

Dr Geissler: I think it all depends on how much effort you are willing to put into this, and I think initially we were looking for a scoring system that is simple to apply, that is readily available and that gives excellent results, and apparently there is no such thing around, and I think if you really want to have excellent predictive values, you need to split the thing up into different variables and everything else. That is probably true.

Dr A. Royse (*Victoria, Australia*): I would just like to emphasize the importance of scoring systems, in the negative. In general with patients you either have normal risk, low risk or high risk, and that is generally quite easy to see clinically. You don't need a computer program. And of these three groups, it is only 'high risk' that actually means very much, because it is the only time you may consider changing something in your treatment.

The second thing I wanted to say pertains to the various types of scoring systems. The scores are based on your experience, with your patients, at your institution, at that time. You cannot transport that to some other place or even to yourselves forward in time. There was a classic illustration of this in your paper, where the Parsonnet scoring system, taken from another country and another time frame, was no longer applicable to you, and I think that it is very important to appreciate the limitations of any scoring system.

Dr Geissler: I think you are absolutely correct, and actually it was the purpose of our study to examine the applicability of these scores in our patient population.