ORIGINAL ARTICLE

Mortality prediction in adult cardiac surgery patients: comparison of two risk stratification models

WK Au 區永谷
MP Sun 沈文萍
KT Lam 林國泰
LC Cheng 鄭力翔
SW Chiu 趙瑞華
SR Das

Objective To assess and compare the two commonly applied models—EuroSCORE and Parsonnet—in our local adult cardiac surgery patients, according to risk factor quantification related to mortality using a risk stratification protocol to assess the quality of cardiac surgical care.

Design Prospective study.

Setting Cardiac surgery centre in a regional hospital in Hong Kong.

Patients All adult patients undergoing coronary artery bypass graft and heart valve surgery at the Grantham Hospital were evaluated prospectively from November 1999 to July 2005.

Main outcome measures In-hospital mortality was the defined end-point. Statistical analyses consisted of observed against expected mortality, Hosmer-Lemeshow goodness-of-fit test for calibration accuracy, and receiver operating characteristic curve for discrimination performance.

Results During the study period, 1247 patients underwent coronary artery bypass graft surgery and 1406 underwent heart valve surgery. Observed mortality rates in these two patient groups were 2.9% and 4.8% respectively. The expected mortality rates as predicted by the EuroSCORE were (mean±standard deviation) 4.0±3.3% and 5.2±3.0% respectively, and by the Parsonnet model were 5.9±4.2% and 7.3±4.4% respectively. EuroSCORE performed better than the Parsonnet model at predicting in-hospital mortality assessed by the Hosmer-Lemeshow goodness-of-fit test. The areas under the receiver operating characteristic curves in coronary artery bypass graft surgery were 0.76 for EuroSCORE and 0.74 for Parsonnet. The receiver operating characteristic curve areas in valve surgery were 0.77 for EuroSCORE and 0.79 for Parsonnet.

Conclusion Despite significant geographic and demographic differences between European and Asian patients, in our local adult cardiac surgery patients, the EuroSCORE performed well with good calibration and discrimination in predicting mortality. There was a tendency for both models to over predict. However, the EuroSCORE can serve as a baseline for the development of a local risk model.

Key words
Heart diseases; Hospital mortality; Prospective studies; Risk assessment; Risk factors

Hong Kong Med J 2007;13:293-7

The Grantham Hospital, Aberdeen, Hong Kong:
Division of Cardiothoracic Surgery
WK Au, FRCS, FHKAM (Surgery)
MP Sun, BSc (RN)
KT Lam, MPhil
LC Cheng, FRCS, FHKAM (Surgery)
SW Chiu, FRCS, FHKAM (Surgery)
Department of Anaesthesia
SR Das, FFARCSI, FHKAM (Anaesthesiology)

Correspondence to: Dr WK Au
E-mail: auwkt@ha.org.hk

Introduction

Analysis of patient outcome has gained increasing importance as public and health authorities demand more data on risk, prognosis, and performance of specific procedures, particularly for resource-intensive operations such as coronary artery bypass graft (CABG) and heart valve surgery. It also helps the patient as well as the family to weigh the risks and benefits so as to clarify expectations.

Crude mortality rate is an easy and readily measurable outcome but it is obvious that simple comparison of postoperative mortality does not reflect the actual quality of service of institutions or individual surgeons, due to wide variations in preoperative patient morbidity.[1,2]

從預測成年心臟外科病人的死亡率以比較兩個風險分層模式

目的 以本地成年心臟外科病人為研究對象，根據死亡率的風險因素，評估和比較兩個常用的風險分層模式（EuroSCORE和Parsonnet）在方法上評估心臟手術的效果。

設計 前瞻性研究。

安排 香港一所地區醫院的心臟外科中心。

患者 由1999年11月至2005年7月在葛量洪醫院進行冠狀動脈搭橋手術和心瓣手術的所有病人。

主要結果測量 住院期間死亡被定為研究終點。預測死亡率及觀察死亡率、Hosmer-Lemeshow擬合優度檢驗（用作校正準確度）、和受試者操作曲線（ROC曲線，用作辨別表現）等統計用以作分析。

結果 研究期間，有1247名病人接受冠狀動脈搭橋手術，1406名病人接受心瓣手術。兩種手術的觀察死亡率分別為2.9%和4.8%，利用EuroSCORE模式預測的死亡率（平均值±標準差）分別為4.0±3.3%和5.2±3.0%，而Parsonnet模式預測的死亡率則分別為5.9±4.2%和7.3±4.4%。經過Hosmer-Lemeshow擬合優度檢驗的評估後，在預測住院死亡率上，EuroSCORE的表現較好。至於冠狀動脈搭橋手術的ROC曲線：EuroSCORE為0.76，Parsonnet為0.74；心瓣手術的ROC曲線：EuroSCORE為0.77，Parsonnet為0.79。

結論 雖然歐洲和亞洲病人有明顯的地理和人口學的差異，但以是次本地成年心臟外科病人的例子上，EuroSCORE模式在預測死亡率方面有較好的校正和辨別表現。兩種模式預測的數字都偏高，然而EuroSCORE模式可作為發展本地風險評估模式的基礎。

Various mortality prediction (risk stratification) models have been developed to correct for differences in the patient population and allow comparison of actual outcome to expected outcome in cardiac surgery.[3-5] The aim of this study was to assess and compare the two commonly applied models—the EuroSCORE and Parsonnet models—in our local adult cardiac surgery patients.[3,4,6]

Methods

All patients undergoing either CABG or heart valve surgery at the Grantham Hospital from November 1999 to July 2005 inclusive were recruited for this study; adults undergoing congenital heart and aortic surgery were excluded. Data were collected prospectively by two designated medical staff according to the criteria and definitions described by the model developers.[4,6] Random cross-checking was performed to minimise interobserver errors. A nurse specialist was assigned for mortality registration and data entry. Two risk stratification models were selected: the EuroSCORE model and the Parsonnet model. The EuroSCORE is one of the most updated risk stratification models available. It was developed with data collected from 130 centres in eight European countries. It recruited around 20 000 patients and was first published in 1999.[3] The Parsonnet model was first published in 1989 and is by far one of the most commonly applied models in adult heart surgery.[4]

The primary outcome of this study was in-hospital mortality, which was defined as either postoperative mortality within the same hospital admission, or within 30 days of surgery.

Statistical analysis

Expected mortality (in percentage terms) was defined as the summation of the calculated mortality scores from the respective models divided by the number of patients. Observed (crude) mortality was the actual number of corresponding deaths divided by the number of patients. The performance of different risk stratification models can be assessed by using calibration and discrimination statistical analysis. Calibration analysis was performed using the Hosmer-Lemeshow goodness-of-fit test, which compares observed and expected occurrences of the outcome (predefined mortality) for each decile by the Chi squared test. The null hypothesis is that the model does not fit; thus the combination of P value greater than 0.05 together with a low C statistic (χ^2) value indicates a good model.[7,8] The receiver operating characteristic (ROC) curve was used for the discrimination analysis. The area under the ROC curve (from 0 to 1.0) correlates with the discriminatory capability of the model. In general, the larger the area under the ROC curve, the better the discriminatory power of that particular model.[8,9] However, the ROC curve value is valid and meaningful only after the model has been shown to calibrate well.

Results

Between November 1999 and July 2005, 2653 consecutive patients who underwent either CABG or heart valve surgery were recruited prospectively. The demographic characteristics of each patient group are shown in Tables 1 and 2 respectively. Table 3 shows the predicted mortality of the two risk stratification models compared with the actual mortality. Calibration assessment of the two risk stratification models was performed, using Hosmer-Lemeshow goodness-of-fit test (Table 4). Discrimination analysis of the two models using ROC curves is shown in Figures 1 and 2, together with areas under the ROC curves for comparison.

Discussion

As patient populations may differ significantly

TABLE 1. Demographic characteristics of coronary artery bypass graft surgery patients (n=1247)

Preoperative risk factor	No. (%)
Male	950 (76)
Hypertension	863 (69)
Morbid obesity	16 (1)
Diabetes mellitus	521 (42)
Ejection fraction 30-49%	258 (21)
Ejection fraction <30%	77 (6)
Aborted death/ventricular fibrillation and ventricular tachycardia	25 (2)
Re-operation	16 (1)
Transmyocardial infarction <48 hours	25 (2)
Congestive heart failure	190 (15)
Preoperative aortic balloon pump	59 (5)
Renal failure	86 (7)
Peripheral vascular disease	96 (8)
Left main stenosis >50%	513 (41)
Carotid stenosis >50%	165 (13)

TABLE 2. Demographic characteristics of heart valve surgery patients (n=1406)

Preoperative risk factor	No. (%)
Male	672 (48)
Hypertension	243 (17)
Diabetes mellitus	127 (9)
LV* ejection fraction 30-49%	269 (19)
LV ejection fraction <30%	35 (2)
Re-operation	288 (20)
Congestive heart failure	1092 (78)
Pulmonary systolic pressure >60 mm Hg	145 (10)
Renal failure	67 (5)
Neurological dysfunction	82 (6)
Tricuspid valve surgery	326 (23)

* LV denotes left ventricule

TABLE 3. Expected and observed mortality of the adult cardiac surgical patients

Surgery	EuroSCORE, mean±SD	Parsonnet model, mean±SD	Observed mortality
Coronary artery bypass graft	4.0±3.3%	5.9±4.2%	2.9%
Heart valve	5.2±3.0%	7.3±4.4%	4.8%

TABLE 4. Hosmer-Lemeshow goodness-of-fit test for the EuroSCORE and Parsonnet models

Model/surgery	C statistic (χ^2) value	P value
EuroSCORE		
Coronary artery bypass graft	2.799	0.903
Heart valve	2.319	0.803
Parsonnet		
Coronary artery bypass graft	5.950	0.653
Heart valve	11.446	0.178

between institutions and countries, a simple comparison of crude mortality is neither an accurate nor feasible means of auditing surgical outcome. In this study, the numbers of patients in the CABG and heart valve surgery groups were roughly the same. However, in the EuroSCORE study, an estimated 65% of the population had CABG procedures and only 35% had valve surgery. A relatively high percentage of our CABG patients had risk factors such as diabetes mellitus (42%), left main coronary disease (41%), whereas fewer were undergoing re-operation (1%) or had morbid obesity (1%). Such risk factor variations were also noted in our heart valve surgery group which had a relatively high incidence of re-operation (20%) and tricuspid valve surgery (23%).[3,10] Differences in the incidence of such risk factors could affect the performance of risk stratification models applied to different geographical and ethnic populations.[10,11] Furthermore, no Asian-based risk stratification model has been developed for regional application. Therefore, before applying western risk stratification models to our local population, we need to verify their applicability, predictability, and accuracy.

In this study, 'in-hospital' mortality was the chosen end-point. Mortality has been the most commonly used outcome in the majority of the risk stratification models, as it is easily defined and easily measured. Some studies showed that postoperative morbidities, length of hospital stay, and patient satisfaction are also important factors in assessing the quality of health service provided. However, some factors have little standardisation and lack a universal definition. As a result accurate documentation of events remains very difficult.[12,13]

In our adult cardiac surgery patients, the results of the calibration and discrimination statistical analyses revealed that the EuroSCORE model performed better than the Parsonnet model

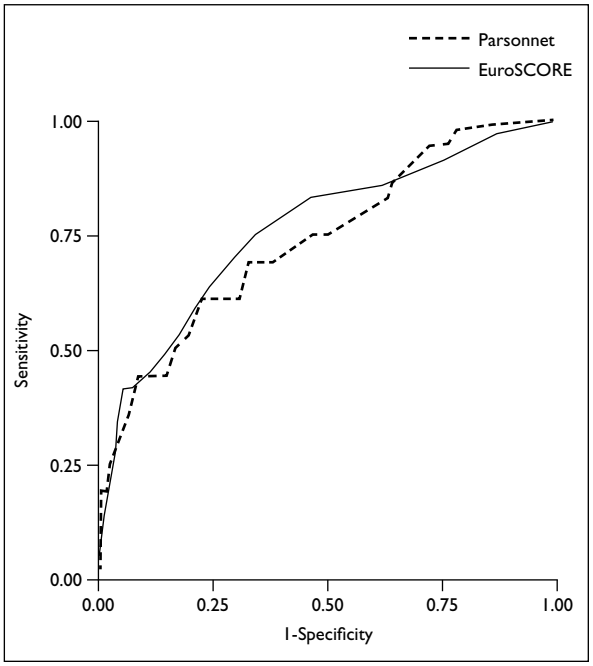

FIG 1. Comparison of receiver operating characteristic (ROC) curves of the EuroSCORE and Parsonnet models in coronary artery bypass graft surgery patients
Area under the ROC curve for EuroSCORE and Parsonnet models were 0.76 (95% confidence interval, 0.68-0.85) and 0.74 (0.65-0.83), respectively

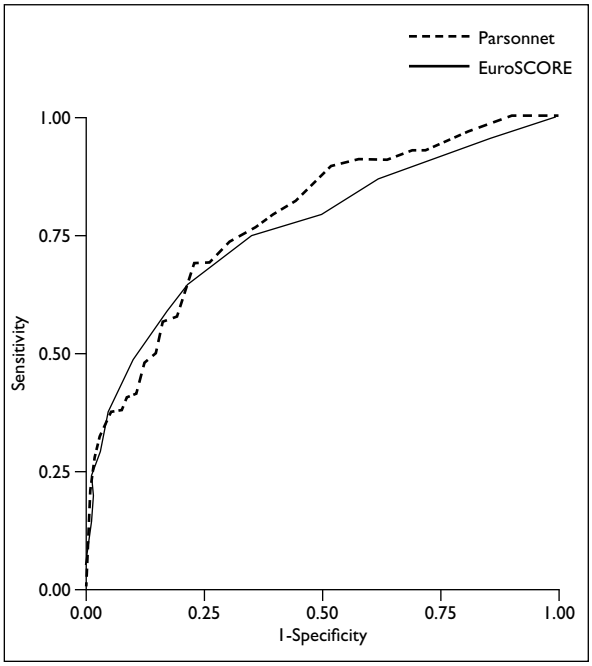

FIG 2. Comparison of receiver operating characteristic (ROC) curves of the EuroSCORE and Parsonnet models in heart valve surgery patients
Area under the ROC curve for EuroSCORE and Parsonnet models were 0.77 (95% confidence interval, 0.70-0.83) and 0.79 (0.73-0.84), respectively

in predicting postoperative mortality. In Table 4, EuroSCORE showed lower *C* statistical values and higher P values indicating good agreement between the observed and expected number of deaths in both our CABG and heart valve surgery patients. Although the area under the ROC curve in the heart valve surgery group was slightly higher in the Parsonnet than EuroSCORE model (0.79 vs 0.77), the calibration performance of the former was nevertheless much better. As mentioned already, ROC curve analysis is valid and meaningful only after the model has been shown to calibrate well in the Hosmer-Lemeshow goodness-of-fit test.[8] In practice, the most commonly used risk stratification models have areas under the ROC curve in the range of 0.70 to 0.85.[12] Therefore, we demonstrated that the EuroSCORE model performed better than the Parsonnet model in terms of postoperative mortality prediction in the Hong Kong Chinese population undergoing cardiac surgery. Perhaps the Parsonnet model performed poorer in this study because it was developed more than 10 years ago and recent surgical and medical advances could weaken its predictive power.

In our CABG group, the observed mortality was 2.9% while the expected mortalities calculated by EuroSCORE and Parsonnet were 4.0% and 5.9% respectively. It is possible that the surgical outcome of our CABG patients is better than that in average European and American institutions as reflected by predicted mortality values. Another contributing factor that could explain the discrepancy might be related to geographical and racial differences, as suggested in other European and Japanese studies.[11,14,15] The difference between observed and expected mortality rates provides an outcome-based measure of quality of care and provides a sound baseline for financial and human resource allocation by health care administrators. Currently, there is no valid model developed for Asian regional use. Therefore, there is a strong need to create a local risk stratification model with even better predictive power. Statistical analysis using logistic regression can identify independent risk factors affecting the in-hospital mortality and a local scoring system can be derived from respective odds ratios.[1,2,16] The Bayesian statistical formula gives additional accuracy for the development of a local risk model. The readers are referred elsewhere for more detailed discussion of statistical methods used in outcome assessment models.[17,18]

Conclusion

Risk stratification models for mortality prediction in adult heart surgery are potent tools. They provide accurate adjusted outcomes for audit purposes

and important information to patients and their relatives about procedural risks. EuroSCORE is easy to apply with clearly defined variables. It has been shown in this series to perform reasonably well in our local Chinese patients undergoing CABG and heart valve surgery, despite having been developed in European populations. However, the ultimate goal is to create a local risk stratification model that can be used for accurate mortality prediction in our own community, whilst taking into account local demographic variables and advances in medical technology. Meanwhile, the EuroSCORE can be used for local audit purpose and serve as the foundation for developing a local risk stratification model.

References

1. Shahian DM, Normand SL, Torchiana DF, et al. Cardiac surgery report cards: comprehensive review and statistical critique. Ann Thorac Surg 2001;72:2155-68.
2. Krumholz HM, Brindis RG, Brush JE, et al. Standards for statistical models used for public reporting of health outcomes: an American Heart Association Scientific Statement from the Quality of Care and Outcomes Research Interdisciplinary Writing Group: cosponsored by the Council on Epidemiology and Prevention and the Stroke Council. Endorsed by the American College of Cardiology Foundation. Circulation 2006;113:456-62.
3. Roques F, Nashef SA, Michel P, et al. Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients. Eur J Cardiothorac Surg 1999;15:816-23.
4. 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:I3-I12.
5. The Society of Thoracic Surgeons. National cardiac surgery database manual for data managers. Minneapolis: Summit medical system; 1993.
6. 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.
7. Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley & Sons; 1989.
8. Martinez-Alario J, Tuesta ID, Plasencia E, Santana M, Mora ML. Mortality predication in cardiac surgery patients: comparative performance of Parsonnet and general severity systems. Circulation 1999;99:2378-82.
9. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839-43.
10. Roques F, Nashef SA, Michel P, et al. Does EuroSCORE work in individual European countries? Eur J Cardiothorac Surg 2000;18:27-30.
11. Pinna-Pintor P, Bobbio M, Colangelo S, et al. Inaccuracy of four coronary surgery risk-adjusted models to predict mortality in individual patients. Eur J Cardiothorac Surg 2002;21:199-204.
12. Higgins TL. Quantifying risk and assessing outcome in cardiac surgery. J Cardiothoracic Vasc Anesth 1998;12:330-40.
13. Nilsson J, Algotsson L, Hoglund P, Luhrs C, Brandt J. EuroSCORE predicts intensive care unit stay and costs of open heart surgery. Ann Thorac Surg 2004;78:1528-34.
14. Geissler HJ, Holzl P, Marohl S, et al. Risk stratification in heart surgery: comparison of six score systems. Eur J Cardiothorac Surg 2000;17:400-6.
15. Kawachi Y, Nakashima A, Toshima Y, Arinaga K, Kawano H. Evaluation of the quality of cardiovascular surgery care using risk stratification analysis according to the EuroSCORE additive model. Circ J 2002;66:145-8.
16. Khuri SF. Quality, advocacy, healthcare policy and the surgeon. Ann Thorac Surg 2002;74:641-9.
17. Edwards FH, Albus RA, Zajtchuk R, et al. Use of a Bayesian statistical model for risk assessment in coronary artery surgery. Ann Thorac Surg 1988;45:437-40.
18. Keogh B, Kinsman R. National adult cardiac surgical database report. The Society of Cardiothoracic Surgeons of Great Britain and Ireland; 1999.