ELSEVIER

European Journal of Cardio-thoracic Surgery 25 (2004) 695–700

EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY

www.elsevier.com/locate/ejcts

Review

EuroSCORE: a systematic review of international performance

A. Gogbashian[a,*], A. Sedrakyan[a], T. Treasure[b]

[a]*The Royal College of Surgeons of England, 35-43 Lincoln's Inn Fields, London WC2A 3PN, UK*
[b]*Department of Cardiothoracic Surgery, Guy's and St Thomas' Hospital NHS Trust, Guy's Hospital, St Thomas Street, London SE1 9RT, UK*

Received 31 December 2003; received in revised form 13 February 2004; accepted 19 February 2004

Summary

The validity of the cardiac surgical scoring system, EuroSCORE, has been assessed by several individual cardiac centres within and outside Europe. We chose to assess the overall international performance by systematic review of peer-reviewed literature. There were six studies meeting our criteria for assessment. Internationally, the evidence is highly suggestive that additive EuroSCORE performance generally over-estimates mortality at lower EuroSCOREs (EuroSCORE ≤ 6) and under-estimates mortality at higher EuroSCOREs (EuroSCORE > 13). The effect of this could have serious misrepresentations for surgeons and hospitals operating on differing case-mixes. We suggest that further studies need to be performed on the logistic EuroSCORE calculation to ascertain whether predictive ability is improved. Overall, however, EuroSCORE is the most rigorously evaluated scoring system in cardiac surgery.

Keywords: Cardiac surgery; EuroSCORE; Mortality; Risk stratification

1. Introduction

EuroSCORE is a prognostic scoring system developed in Europe for patients undergoing cardiac surgery [1]. Earlier scoring systems for predicting mortality in cardiac surgery include Parsonnet, The Cleveland Clinic coronary scoring system and the UK Society of Cardiothoracic Surgeons Risk Score [2–4]. These older scoring systems were developed from a single database of patients, the first two institutional and the third from a national database. There has been debate as to whether they can be generalised to other practices or other countries. Parsonnet was the founding father of systematic risk stratification [2] and the first to demonstrate that his method could be applied to European practice; Nashef and colleagues validated the method in Manchester, UK [5]. The principle was quickly adopted in UK practice [6] and incorporated in risk-adjusted CUSUM displays [7]. However, there has also been a question of whether the earlier systems were adequately objective [8] prompting a European scoring system based on explicit objective criteria.

1.1. The EuroSCORE

There was therefore an incentive to develop a more robust and objective scoring system to cover a wider population of cardiac surgical patients. EuroSCORE was derived from an European database of nearly 20,000 consecutive patients from 128 hospitals in eight European countries. Information on 97 risk factors was collected pre-operatively in all the patients. These risk factors were then compared to patient outcomes (survival or death). By means of logistic regression calculations, those risk factors that were robust in predicting mortality became part of the EuroSCORE calculation [9]. Two different scoring systems exist for the EuroSCORE. One is known as the additive model—a score that can be calculated by simple arithmetic—and the other known as the logistic model. The logistic model is more extensive and requires a computer to derive a score [10]. Since the EuroSCORE has been developed, it has been tested on several local populations around the world, including Europe, Japan, Scandinavia and the USA [11–15].

1.2. Objective

There have been several individual centre studies as well as regional studies examining the effectiveness and validity of the EuroSCORE at a local level. We desired to undertake

* Corresponding author. Tel./fax: +44-208-731-6176.
E-mail address: andrew@cardiacforum.com (A. Gogbashian).

1010-7940/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ejcts.2004.02.022

a review of overall international performance of EuroSCORE by assessing the available literature.

2. Materials and methods

2.1. Search strategy

Published studies, using the EuroSCORE as a means of risk stratification for patients undergoing cardiac surgery, were identified by searching Medline between 1/1966 and 8/2003; Embase between weeks 01/1980 and 34/2003; Current Contents between weeks 1/1993 and 35/2003. Search terms used for finding the articles were 'Euroscore', 'Cardiac data' and (cardi* or surg* or score*). Both English language and foreign language journals were searched. The references cited by these articles were then searched for further articles related to the EuroSCORE. Two of the authors searched independently and all authors participated in discussion of discrepancies.

2.2. Inclusion criteria

Only papers that included surgery on adult patients and reported findings based on prospectively collected data were included. Both the logistic calculation and the additive calculations of the EuroSCORE were accepted. Data were accepted regardless of the type of cardiac operation performed. Papers that separated patients into low and high-risk groups were accepted.

2.3. Exclusion criteria

Data from the original database by which the EuroSCORE was derived were excluded from the study. Also excluded were papers with identical or overlapping patient samples; studies that did not include a separation of their dataset into differing EuroSCORE risk groups; and articles from which we could not derive observed and predicted mortality.

2.4. Data extraction

Data were extracted directly from the text, tables and graphs from included papers. We attempted to get additional information from the authors by electronic mail to addresses cited in papers. Where the studies did not provide the full data required and the author did not provide the data by further correspondence we were unable to include them. Further data in addition to those provided in the original article were provided by Sergeant et al. [16]. In the study by Bridgewater et al. [17] data were extracted by reading off the axes of these graphs.

3. Results

3.1. Overview of included and excluded papers

Studies meeting the criteria can be found in Table 1.

All six studies suitable for inclusion used the additive EuroSCORE calculation.

Table 2 shows the articles excluded from the systematic review and the reasons.

3.2. Observed and expected mortality

Table 3 summarises the observed and expected mortality of operations at various EuroSCORE sub-groups.

Table 1
Summary of reviewed studies

Study	Country of data origin	Year of publication	Total number of patients	EuroSCORE calculation
Kawachi et al. [15]	Japan	2001	803	Additive
Sergeant et al. [16]	Belgium	2001	2051	Additive
Stoica et al. [18]	France	2002	1543	Additive
Karabulut et al. [19]	Turkey	2003	1123	Additive
Al-Ruzzeh et al. [20]	UK	2003	1907	Additive
Bridgewater et al. [17]	UK	2003	8572	Additive

Table 2
Summary of excluded studies

Study	Country of data origin	Year of publication	Reasons for exclusion
Nashef et al. [1]	Europe	1999	Original database used for EuroSCORE derivation
Pitkänen et al. [14]	Finland	2000	Inadequate predicted mortality data
Calafiore et al. [21]	Italy	2003	No separation into low and high-risk groups
Asimakopoulos et al. [22]	UK	2003	Unable to read data adequately from graph axes

Table 3
Results of observed and expected mortality

Study	Total number (*n*)	EuroSCORE group	Number patients (*n*)	Observed mortality (*n*)	Observed mortality (%)	Expected mortality (%)	(Observed − expected mortality) (%)
Kawachi et al. [15]	803	0–2	145	0	0.0	1.4	−1.4
		3–5	346	6	1.5	4.0	−2.5
		6–8	193	14	6.8	6.7	0.1
		9–11	73	9	11.0	9.7	1.3
		12 +	46	8	21.0	13.0	8.0
Sergeant et al. [16]	2051	0	191	0	0.0	0.0	0.0
		1	133	0	0.0	1.0	−1.0
		2	225	2	0.9	2.0	−1.1
		3	245	4	1.6	3.0	−1.4
		4	209	1	0.5	4.0	−3.5
		5	234	6	2.6	5.0	−2.4
		6	221	7	3.2	6.0	−2.8
		7	162	5	3.1	7.0	−3.9
		8	126	4	3.2	8.0	−4.8
		9	97	11	11.3	9.0	2.3
		10	64	7	10.9	10.0	0.9
		11	50	3	6.0	11.0	−5.0
		12	31	8	25.8	12.0	13.8
		13	15	5	33.3	13.0	20.3
		14	21	6	28.6	14.0	14.6
		15	7	3	42.9	15.0	27.9
		16	8	4	50.0	16.0	34.0
		17	2	0	0.0	17.0	−17.0
		18	3	2	66.7	18.0	48.7
		19	1	1	100.0	19.0	81.0
		20	4	2	50.0	20.0	30.0
		22	2	0	0.0	22.0	−22.0
Stoica et al. [18]	1540	0–2	420	0	0.0	1.0	−1.0
		3–5	640	12	1.9	2.2	−0.4
		6 +	480	52	10.8	7.4	3.4
Karabulut et al. [19]	1123	0–2	446	5	1.1	1.2	−0.1
		3–5	459	8	3.7	3.8	−0.2
		6 +	218	13	1.2	8.4	−7.2
Al-Ruzzeh et al. [20]	1907	0–2.49	1061	6	0.6	0.9	−0.3
		2.5–4.99	478	5	1.0	3.4	−2.4
		5–9.99	344	12	3.5	6.0	−2.5
		10–19.99	24	3	12.5	10.4	2.1
		20 +	0	0	0.0	0.0	0.0
Bridgewater et al. [17]	8572	0	1200	6	0.5	0.0	0.5
		1	1400	10	0.7	1.0	−0.3
		2	1300	10	0.8	2.0	−1.2
		3	1250	14	1.1	3.0	−1.9
		4	1100	15	1.4	4.0	−2.6
		5	800	16	2.0	5.0	−3.0
		6	500	12	2.4	6.0	−3.6
		7	300	14	4.7	7.0	−2.3
		8	200	14	7.0	8.0	−1.0
		9	100	4	4.0	9.0	−5.0
		10	25	3	12.0	10.0	2.0
		11	25	8	32.0	11.0	21.0
		12	25	3	12.0	12.0	0.0
		13	10	1	10.0	13.0	−3.0
		14	10	4	40.0	14.0	26.0
		15	10	5	50.0	15.0	35.0
		16	10	3	30.0	16.0	14.0
		17	10	2	20.0	17.0	3.0
		19	10	1	10.0	19.0	−9.0

Expected mortality (%) is the predicted EuroSCORE.

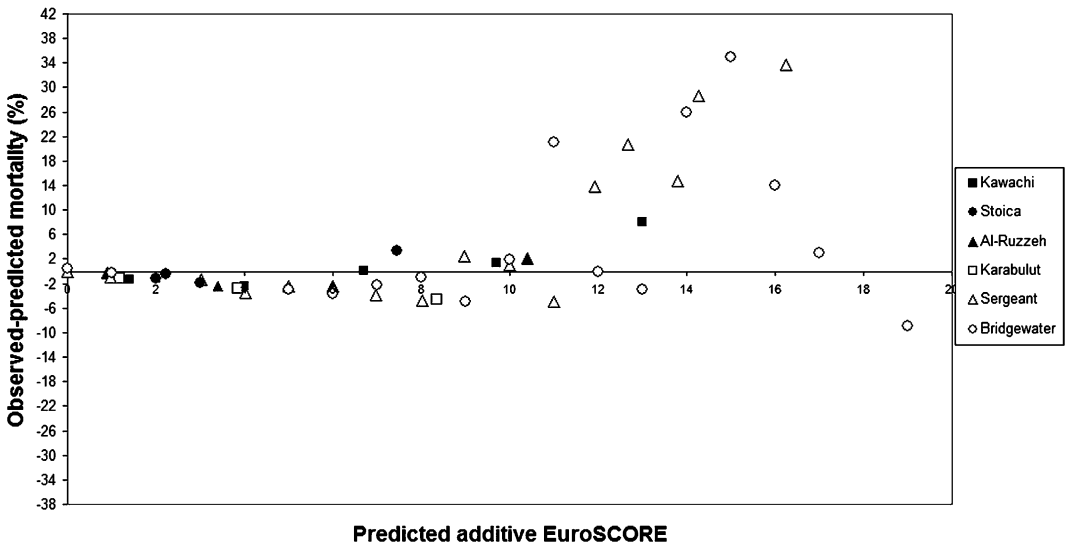

Fig. 1. Predicted additive EuroSCORE plotted against observed − predicted mortality %.

The method of assimilating patients into EuroSCORE groups was particular to each study.

Only the article by Bridgewater et al. [17] contained data for ungrouped EuroSCOREs. Sergeant et al. [16] provided individual EuroSCORE data by further correspondence. All other authors assembled data within EuroSCORE groups, however, they did provide the groupings-predicted mortality as a percentage—in other words, the average EuroSCORE for the group. Observed mortality, as a percentage, was calculated from either the number of deaths quoted or a direct statement of percentage mortality from the studies.

Fig. 1 shows expected mortality (predicted EuroSCORE) plotted against observed − expected mortality% for each of the studies. If there is to be a good 'match' between observed and predicted mortality—an indicator that the EuroSCORE is predicting mortality accurately—the majority of points will cluster around 0 on the *y*-axis. Fig. 1 suggests (in this depiction there are no confidence intervals) that at EuroSCORE ≤ 6 the additive EuroSCORE overestimates risk (expected is greater than observed) in several of studies. At EuroSCORE ≥ 10, the overall appearance from the figure is that additive EuroSCORE underestimates risk.

Table 4 summarises combined data from all six studies. Data were combined to provide an overall picture of EuroSCORE performance. At higher EuroSCORE percentages there are generally fewer numbers of patients in each individual study. We therefore started at the high-risk end creating groups of 100 or more patients, thus narrowing the confidence intervals (see Fig. 2). Data at the very high end of EuroSCORE had to be merged into an EuroSCORE group of 11–24.

Fig. 2 displays predicted mortality (EuroSCORE) on the *x*-axis against observed mortality (*y*-axis) for the combined data, with confidence intervals. The line of identity is where the points would fall for perfect prediction. This figure shows the significant over-estimation of additive

Table 4
Combined data by weighted mean

EuroSCORE group	Total number of patients	Weighted mean observed mortality%	Weighted mean predicted mortality%	95% Confidence interval
11–24	300	23.0	13.3	18.4–28.2
10	186	11.3	9.9	7.1–16.7
9	197	7.6	9.0	4.3–12.2
8	544	4.8	8.1	3.1–6.9
7	1135	7.4	7.1	6–9.2
6	1065	2.9	6.0	2–4.1
5	1034	2.1	5.0	1.3–3.2
4	2114	1.2	4.0	0.8–1.9
3	1973	1.2	3.1	0.7–1.7
2	2165	1.1	2.1	0.7–1.6
1	3605	0.6	1.0	0.4–0.9
0	1391	0.4	0.0	0.2–0.9

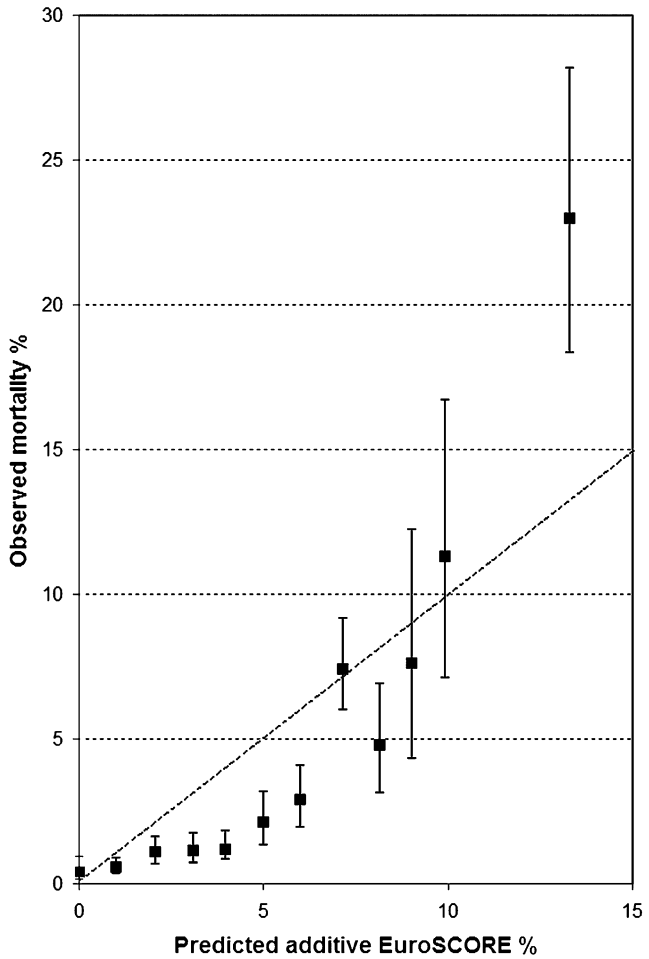

Fig. 2. Overall international performance: predicted additive EuroSCORE against observed mortality %.

EuroSCORE at lower predictions: EuroSCORE ≤ 6. Additive EuroSCORE seems to overestimate risk above 13.

4. Discussion

There were large numbers of patients in all the articles reviewed at lower additive EuroSCORE predictions. The conclusion that EuroSCORE systematically over-estimates risk for these patients is therefore a robust finding which applies for all the studies representing five countries. It is therefore likely to be a universally true finding. In order to assess additive EuroSCORE performance at higher scores it was necessary to combine the data to produce significant numbers in order to produce a meaningful result. The conclusion is less certain but appears to be valid. Both Bridgewater et al. [17] and Sergeant et al. [16] have made this observation within their own data sets. What we have added is that this holds in all instances reported and when the data are pooled it becomes more certain. The consequence is that additive EuroSCORE is likely to be a little forgiving of high volume practice in low-risk patients. However, it has the effect of penalising the surgeon taking on high-risk cases. This is likely to be unimportant in a large mixed practice but will reflect adversely on the surgeon with a lower volume of routine cases and a relatively high proportion of unstable and higher risk patients. This was exactly the reason for Parsonnet introducing risk adjustment. The high-volume private USA practices of the 1980s reported operative mortality rates approaching zero which could not be matched by units with adverse case-mix or disadvantaged populations. The consequence is that the most deserving of patients, the ones where the difference between prospects without and with surgery are the greatest, may be deprived of the chance of surgery as surgeons protect their figures and their reputations.

Bridgewater's suggestion [17] is only to make comparison between surgeons and units for performance monitoring purposes at the lower risk end of the spectrum, the pros and cons of which he explores thoughtfully.

Michel et al. [10] suggest that the logistic EuroSCORE model will resolve this, a contention we have been unable to explore further because we have found insufficient reported data on which to test this. However, they base their assertion on data analysed by ROC curves. We agree with Sergeant et al. [16] that the ROC method merely balances the prediction errors at the low and high-risk ends. We prefer our more intuitively obvious and simpler display of how the risk model performs. ROC analysis is more complicated and more appropriate for the trade-off of sensitivity and specificity in setting a single threshold for a diagnostic test. In our view it may be misleading to regard EuroSCORE as a diagnostic test of death. Furthermore their own analysis shows very near identical ROC curves for additive and logistic EuroSCORE [10].

It may be argued that centres publishing their results within peer-reviewed literature may have overly favourable results, thus providing publication bias. However, centres that have published favourable mortality compared to predicted mortality by additive EuroSCORE within the same publication publish less favourable mortality rates at the higher predictions. This effect is reproducible in several of the studies and with large numbers of patients. We believe it is more likely that it is an inherent feature of additive EuroSCORE than publication bias.

In spite of these flaws EuroSCORE is overall the best-established and validated risk model for contemporary practice. However, it is important to continue to apply professional judgement and common sense in the interpretation of surgical results and to avoid making inappropriate comparisons that disadvantage both patient and surgeon.

4.1. Consideration for further research

From the systematic review performed there were no individual centre studies that calculated the logistic EuroSCORE from their patient data. New studies directly focusing on the logistic EuroSCORE could be started,

however, former data from previous studies could still be utilised, as the input data for the required calculation is the same as for additive EuroSCORE. A systematic review could then be performed on the logistic EuroSCORE.

References

[1] 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(1):9–13.

[2] 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;701(Suppl):13–112.

[3] 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; 267:2344–8.

[4] Keogh BE, Kinsman R. The society of cardiothoracic surgeons of Great Britain and Ireland: National adult cardiac surgical database report 2000–2001

[5] Nashef SA, 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;305:1066–7.

[6] Treasure T. Risks and results of surgery. Br Heart J 1995;74:11–12.

[7] Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life-adjusted display. Lancet 1997;350(9085):1128–30.

[8] Bridgewater B, Neve H, Moat N, Hooper T, Jones M. Predicting operative risk for coronary artery surgery in the United Kingdom: a comparison of various risk prediction algorithms. Heart 1998;79:350–5.

[9] Roques F, Nashef SA, Michel P, Gauducheau E, de Vincentiis C, Baudet E, Cortina J, David M, Faichney A, Gabrielle F, Gams E, Harjula A, Jones MT, Pintor PP, 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(6):816–22. discussion 822–3.

[10] Michel P, Roques F, Nashef SA, EuroSCORE Project Group. Logistic or additive EuroSCORE for high-risk patients? Eur J Cardiothorac Surg 2003;23(5):684–7.

[11] Roques F, Nashef SA, Michel P, Pintor PP, David M, Baudet E, The EuroSCORE Study Group. Does EuroSCORE work in individual European countries? Eur J Cardiothorac Surg 2000;18(1):27–30.

[12] Nashef SA, Roques F, Hammill BG, Peterson ED, Michel P, Grover FL, Wyse RK, Ferguson TB, EuroSCORE Project Group. Validation of European system for cardiac operative risk evaluation (EuroSCORE) in North American cardiac surgery. Eur J Cardiothorac Surg 2002;22(1):101–5.

[13] 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(2):145–8.

[14] Pitkanen O, Niskanen M, Rehnberg S, Hippelainen 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(6):703–10.

[15] Kawachi Y, Nakashima A, Toshima Y, Arinaga K, Kawano H. Risk stratification analysis of operative mortality in heart and thoracic aorta surgery: comparison between Parsonnet and EuroSCORE additive model. Eur J Cardiothorac Surg 2001;20(5):961–6.

[16] Sergeant P, de Worm E, Meyns B. Single centre, single domain validation of the EuroSCORE on a consecutive sample of primary and repeat CABG. Eur J Cardiothorac Surg 2001;20(6): 1176–82.

[17] Bridgewater B, Grayson AD, Jackson M, Brooks N, Grotte GJ, Keenan DJ, Millner R, Fabri BM, Jones M, North West Quality Improvement Programme in Cardiac Interventions. Surgeon specific mortality in adult cardiac surgery: comparison between crude and risk stratified data. Br Med J 2003;327(7405):13–17.

[18] 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(1):41–6.

[19] Karabulut H, Toraman F, Alhan C, Camur G, Evrenkaya S, Dagdelen S, Tarcan S. EuroSCORE overestimates the cardiac operative risk. Cardiovasc Surg 2003;11(4):295–8.

[20] Al-Ruzzeh A, Asimakopoulos G, Ambler G, Omar R, Hasan R, Fabri B, El-Gamel A, DeSouza A, Zamvar V, Griffin S, Keenan D, Trivedi U, Pullan M, Cale A, Cowen M, Taylor K, Amrani M. Validation of four different risk stratification systems in patients undergoing off-pump coronary artery bypass surgery: a UK multicentre analysis of 2223 patients. Heart 2003;89(4):432–5.

[21] Calafiore AM, Di Mauro M, Canosa C, Di Giammarco G, Iaco AL, Contini M. Early and late outcome of myocardial revascularization with and without cardiopulmonary bypass in high risk patients (EuroSCORE $\geq$ 6). Eur J Cardiothorac Surg 2003;23(3): 360–7.

[22] Asimakopoulos G, Al-Ruzzeh S, Ambler G, Omar RZ, Punjabi P, Amrani M, Taylor KM. An evaluation of existing risk stratification models as a tool for comparison of surgical performances for coronary artery bypass grafting between institutions. Eur J Cardiothorac Surg 2003;23:935–42.