Tadahiro Ikeoka, Dimas; Fernandes, Viviane Aparecida; Gebara, Otávio; Teixeira Garcia, José Carlos; Melo de Barros e Silva, Pedro Gabriel; Jamus Rodrigues, Marcelo; Furlan, Valter; do Amaral Baruzzi, Antônio Cláudio

Evaluation of the Society of Thoracic Surgeons score system for isolated coronary bypass graft surgery in a Brazilian population

Revista Brasileira de Cirurgia Cardiovascular/Brazilian Journal of Cardiovascular Surgery, vol. 29, núm. 1, enero-marzo, 2014, pp. 51-58

Sociedade Brasileira de Cirurgia Cardiovascular
São José do Rio Preto, Brasil

Available in: http://www.redalyc.org/articulo.oa?id=398941892011
Evaluation of the Society of Thoracic Surgeons score system for isolated coronary bypass graft surgery in a Brazilian population

Avaliação dos escores da Society of Thoracic Surgeons para cirurgia de revascularização miocárdica isolada em uma população brasileira

Dimas Tadahiro Ikeoka¹, MD, PhD; Viviane Aparecida Fernandes¹; Otávio Gebara¹, MD, PhD; José Carlos Teixeira Garcia¹, MD; Pedro Gabriel Melo de Barros e Silva¹, MD; Marcelo Jamus Rodrigues¹, MD; Valter Furlan¹, MD; Antônio Cláudio do Amaral Baruzzi, MD, PhD¹

Abstract

Objective: Report the experience with the Society of Thoracic Surgeons scoring system in a Brazilian population submitted to isolated coronary artery bypass graft surgery.

Methods: Data were collected from January-2010 to December-2011, and analyzed to determine the performance of the Society of Thoracic Surgeons scoring system on the determination of postoperative mortality and morbidity, using the method of the receiver operating characteristic curve as well as the Hosmer-Lemeshow and the Chi-square goodness of fit tests. From the 1083 cardiac surgeries performed during the study period 659 represented coronary artery bypass graft procedures which are included in the present analysis. Mean age was 61.4 years and 77% were men.

Results: Goodness of fit tests have shown good calibration indexes both for mortality ($X^2=6.78$, $P=0.56$) and general morbidity ($X^2=6.69$, $P=0.57$). Analysis of area under the ROC-curve (AUC) demonstrated a good performance to detect the risk of death (AUC 0.76; $P<0.001$), renal failure (AUC 0.79; $P<0.001$), prolonged ventilation (AUC 0.80; $P<0.001$), reoperation (AUC 0.76; $P<0.001$) and major morbidity (AUC 0.75; $P<0.001$) which represents the combination of the assessed postoperative complications. STS scoring system did not present comparable results for short term hospital stay, prolonged length of hospital stay and could not be properly tested for stroke and wound infection.

Conclusion: Society of Thoracic Surgeons scoring system presented a good calibration and discrimination in our population to predict postoperative mortality and the majority of the harmful events following coronary artery bypass graft surgery. Analysis of larger samples might be needed to further validate the use of the score system in Brazilian populations.

Descriptors: Risk Management. Cardiovascular Surgical Procedures. Myocardial Revascularization. Postoperative Complications. ROC Curve.
INTRODUCTION

Coronary artery bypass graft (CABG) surgery has been a widely used modality of treatment for coronary artery disease (CAD) for the last few decades. Postoperative outcomes of CABG have progressively improved as a consequence of new technologies, surgical techniques, postoperative care, but also because of better control of specific risk factors, before, during and after surgery [1]. The burden of publications describing risk factors for a worse outcome evolved with the use of large sets of data collected from patients submitted to surgery over distinct periods of time. One important example to be cited is the European System for Cardiac Operative Risk Evaluation (EuroSCORE), a method largely used by many surgeons and cardiologists around the world, to assess the risk of death following cardiac surgery [2]. In the late eighties, the Society of Thoracic Surgeons (STS) also started a visionary effort of collecting data from patients submitted to cardiothoracic surgery all over the United States of America which culminated with the development of the STS National Adult Cardiac Surgery Database, a databank that accounts for more than four and a half millions of surgically treated patients [3,4]. Using the information recorded in that bank, accurate risk scores for morbidity and mortality were developed with help of multiple logistic regression models [5,6].

For the North-American hospitals and surgical facilities who take part at the STS databank, uploading their detailed results is compensated by three-monthly structured performance reports produced and sent back by the society, along with a comprehensive comparison of their results with the entire databank [6]. For those medical professionals around the world who are not STS associated, an online calculator is available at their institutional website [7]. This internet tool enables cardiologists and surgeons all over the world to determine the risk scores for individual patients and prompts their teams for better control of risk factors leading to improved postoperative results.

The present report aims to determine how finely the STS score system could predict complications and mortality in a set of consecutive patients of a private hospital in Brazil, considering that ethnic and socioeconomic characteristics might substantially differ from that of the North-American databank. This is therefore a preliminary analysis of adequacy and calibration of the STS methodology in a South-American population.
METHODS

Study design
This is a single center and observational study, performed on consecutive patients who underwent coronary artery bypass surgery at the TotalCor Hospital, located in São Paulo, Brazil. The study was submitted to the local ethics committee that works in consonance with the Declaration of Helsinki, and approved as presented herein. Written and signed informed consent was waived, since the protocol consisted in a retrospective analysis of an institutional databank. Ethical principles for medical research were entirely assured and no risks were added to that associated with the surgical treatment, as acknowledged by the Ethics Committee (protocol number: 461, approved at December, 28th 2011).

Data management
About two hundred distinct parameters of demographic, clinical and laboratorial nature are routinely collected for each patient who had undergone cardiac surgery in our hospital by a team of quality managers that feed an institutional databank of cardiac surgery on a daily basis. Data collection is performed ahead of the procedure, but also along the hospital stay and after hospital discharge, by telephone in the late postoperative period. Information from this institutional databank is routinely used to detect possible flaws in the treatment and to plan strategies to improve medical practice. Based on these data, STS scores are calculated online in the website as previously cited for each patient and stored as well in the databank. For the purpose of the present investigation, an electronic data-sheet with the relevant variables as extracted from the databank was generated and assessed using the methods described below. Table 1 shows a summary of the baseline clinical variables that were considered relevant for the study.

The website calculator returns a list of estimated risk rates for nine distinct endpoints during the postoperative period, which are defined as follows: 1) operative mortality: death during the in-hospital stay following surgery, regardless of timing, or within 30 days of surgery; 2) permanent stroke (cerebrovascular accident): a central neurologic deficit persisting longer than 72 hours; 3) renal failure: requirement for dialysis or an increase of the serum creatinine to more than 2.0 mg/dL or double the most recent preoperative creatinine level; 4) prolonged mechanical ventilation (longer than 24 hours); 5) deep sternal wound infection (mediastinitis); 6) reoperation for any cause; 7) major morbidity or mortality that include any of the above mentioned events; 8) prolonged postoperative length of stay (PLOS): length of stay (LOS) longer than 14 days; and 9) short postoperative LOS (SLOS): LOS shorter than 6 days with patient alive at discharge. We assessed all nine endpoints as cited for the studied population. A detailed description of the methods used to develop the mathematical models and to calculate the risk rates were previously published in detail elsewhere [6]. We did not register any loss to follow-up.

Statistics
Continuous variables are shown as mean (standard deviation) if normally distributed or median (25-75 percentile) in the cases they don’t fit normality; categorical variables are displayed as absolute number and percentages. The accuracy (sensitivity and specificity) of the STS scores was tested in our population for each individual endpoint using the method of receiver operating characteristic (ROC) curve as described elsewhere [8]. In brief, sensitivity is plotted against “one minus specificity” (1 – specificity) for each value of a specific prognostic score. Area under the curve (AUC) is than calculated and statistically compared with a baseline AUC of 0.50 that indicates prediction no better than chance, and is represented by a diagonal line crossing the graphic area.

The larger is the AUC (closer to 1.0), the higher is considered the capability of the method to predict outcomes. We considered AUC above 0.70 as the limit for adequate discrimination in our analysis. Endpoints that reached a low number of events (5 or less) were excluded from the analysis given the significantly high probability of methodological errors with low number of events. Adequacy between expected and observed endpoints for distinct quintiles of risk was additionally assessed using goodness of fit Chi-square test and the Hosmer-Lemeshow by logistic regression method. Both methods are based on the comparison of expected versus observed events using a Chi-square distribution and considering significance when the descriptive values of P are above the specified value, being in this case 0.05. For the remaining statistics, descriptive levels of P below 0.05 were considered as significant for any two-tailed tests.

RESULTS

From January 2010 to December 2011, one thousand eighty three (1083) patients underwent cardiothoracic surgery in the TotalCor Hospital. The mean age of that population was 61.4 years and 77% were men. Six hundred fifty nine (659) patients underwent isolated CABG surgery. The remaining subjects underwent valve replacement surgery (221 patients), aortic dissection surgery (30 patients), combined – aortic replacement plus CABG – procedures (133 patients), distinct procedures for congenital heart diseases and other cardiac interventions (40 patients). Overall mortality rate was 4.3% (47/1083), whereas in the isolated CABG group the observed mortality was 2.3% (15/659). Overall clinical and demographic characteristics of the CABG population are presented in Table 1. For those patients in this population where a comitant cardiac illness (e.g.: mitral regurgitation and aortic stenosis) was diagnosed, it was con-
sidered mild or of low risk and therefore not treated, at the discretion of the assistant surgeon.

Analysis of sensitivity and specificity – Figure 1 shows the observed mortality in comparison to the mean expected mortality rate as calculated by the STS scoring system. Population was distributed by quintiles of expected events and according to the NCD definitions. General morbidity is displayed similarly in Figure 2. Goodness of fit Chi-Square tests were performed for the end points total mortality and general morbidity and have shown no differences between the expected and the observed mortality (X²=6.78, P=0.56) or morbidity (X²=6.69, P=0.57). In addition, Hosmer-Leme-show goodness of fit tests were performed and contingency tables for respectively mortality and morbidity are presented (Tables 2 and 3).

Fig. 1 – Observed and expected (mean) mortality rates divided by quintiles of expected mortality, as calculated by the STS scoring system. Goodness of fit Chi-square test has shown that the pattern of distribution of expected and observed mortality did not significantly differ in the study population (X²=1.06, P=0.90)

Fig. 2 – Observed (black columns) and expected (mean, white columns) general morbidity rates distributed according to the quintiles of expected mortality, as calculated by the STS scoring system. Goodness of fit Chi-square test has shown that the pattern of distribution of expected and observed morbidity rates did not significantly differ in the study population (X²=6.77, P=0.15)

| Variable                                      | Frequency       |
|-----------------------------------------------|-----------------|
| Patient age (years)                           | 61.2            |
| Female                                        | 24%             |
| Race (percent Caucasian)                      | 95%             |
| Obese or morbid obesity                       | 26%             |
| Diabetes                                      | 42%             |
| Last creatinine level preoperatively (median; 25 - 75 percentile – mg/dL) | 1.0; 0.9 – 1.2 |
| Previous cardiac surgery                      | 2.4%            |
| Cerebrovascular disease                       | 1.8%            |
| Myocardial infarction (< 21 days prior to surgery) | 45%             |
| ST elevation or NST elevation on admission    | 43.3%           |
| Arrhythmia                                    | 3.0%            |
| Coronary vessels (three)                      | 91.5%           |
| Left main disease > or = 50%                  | 14.6%           |
| Ejection fraction < 40%                       | 4.2%            |
| Aortic stenosis                               | 10%             |
| Mitral insufficiency (mild)                   | 30%             |
| Status of the procedure (urgent or emergent)  | 40%             |
Ikeoka DT, et al. - Evaluation of the Society of Thoracic Surgeons score system for isolated coronary bypass graft surgery in a Brazilian population

Table 2. Hosmer-Lemeshow contingency table for goodness of fit as determined for mortality in isolated CABG patients. Calculated Chi-square = 6.89 P=0.55.

| Deciles of risk | Survival | Expected | Death | Expected | Total |
|-----------------|----------|----------|-------|----------|-------|
| 1st             | 1        | 0.65     | 1     | 0.65     | 41    |
| 2nd             | 42       | 0.71     | 0     | 0.71     | 42    |
| 3rd             | 56       | 0.97     | 0     | 0.97     | 56    |
| 4th             | 46       | 0.83     | 0     | 0.83     | 46    |
| 5th             | 46       | 0.88     | 1     | 0.88     | 47    |
| 6th             | 48       | 0.93     | 0     | 0.93     | 48    |
| 7th             | 47       | 1.02     | 2     | 1.02     | 49    |
| 8th             | 46       | 1.10     | 1     | 1.10     | 47    |
| 9th             | 45       | 1.32     | 3     | 1.32     | 48    |
| 10th            | 61       | 5.62     | 6     | 5.62     | 67    |

Table 3. Hosmer-Lemeshow contingency table for goodness of fit as determined for general morbidity in isolated CABG patients. Calculated Chi-square = 6.69 P=0.57.

| Deciles of risk | Event free | Any event | Total |
|-----------------|------------|-----------|-------|
| 1st             | 40         | 2         | 49    |
| 2nd             | 42         | 2         | 50    |
| 3rd             | 44         | 5         | 49    |
| 4th             | 47         | 1         | 48    |
| 5th             | 44         | 7         | 51    |
| 6th             | 44         | 4         | 48    |
| 7th             | 40         | 9         | 49    |
| 8th             | 40         | 10        | 50    |
| 9th             | 38         | 11        | 49    |
| 10th            | 23         | 25        | 48    |

Fig. 3 – ROC-curves showing relation between sensitivity and 1-specificity for STS scoring system for seven post-operative outcomes measured: A) mortality; B) morbidity; C) renal failure; D) prolonged length of hospital stay; E) reoperation; F) prolonged mechanical ventilation and G) short length of hospital stay
ROC-curves describing relation between sensitivity versus 1-specificity for seven, out of nine investigated endpoints, and the calculated area under the curve (AUC) for each measured complication are displayed in Figure 3. The estimated scores of risk have shown to accurately predict the occurrence of renal failure (AUC 0.79; \( P<0.001 \)), prolonged ventilation (AUC 0.80; \( P<0.001 \)), reoperation (AUC 0.76; \( P<0.001 \)), total morbidity (AUC 0.75; \( P<0.001 \)) and mortality (AUC 0.76; \( P<0.001 \)). The use of calculated scores could not appropriately predict the occurrence of shortened length of stay (AUC 0.57; \( P=0.47 \)), whereas permanent stroke and deep wound infection were not analyzed due to a low number of occurrences in the studied population. Although considered significant by the hypothesis test, prolonged length of hospital stay (AUC 0.68; \( P<0.001 \)) cannot be considered as well predicted by the score due to an AUC bellow the pre-specified limit of 0.70.

**DISCUSSION**

The use of the STS methodology is currently disseminated throughout the North-American hospitals, helping to improve the quality of cardiac surgery, as demonstrated by distinct publications [9-11]. Recently, STS databank subscription was opened for candidates outside the USA, and the first institution to participate as an international member was the TotalCor Hospital, a one-hundred beds facility, dedicated to the management of patients with cardiovascular diseases, located in Sao Paulo, Brazil.

A couple of years before the agreement with STS Databank the hospital staff started to routinely use the online calculator at the STS-website for the determinations of the risk for a panel of possible complications for each individual patient. In parallel, risk scores and clinical, surgical and postoperative information were systematically recorded in an institutional database.

Our study assessed these data and indicated that STS scoring system could be able to detect the risk of postoperative complications in our population. From the best of our knowledge it is the first time that the complete set of risk prediction scores is tested in a Brazilian sample and, considering the presented data, it is likely that STS methodology is useful as a tool to safely predict postoperative outcomes after CABG, although further evaluation on larger population samples are needed. Previously, other initiatives have tested individual aspects of the method for one single complication, namely wound sternal infection [12,13], but never in a similar number of subjects.

Risk assessment of adverse outcomes using prediction models has been used in many distinct clinical situations in recent years, with remarkable applications in the prevention of postoperative complications [14-16]. Development these models involves gathering of data in large multicentric databases and meticulous mathematical analysis with help of multiple logistic regression statistics [5,6].

The resulting score systems are useful to determine the risk of complications in populations that share similar risk profiles. Additionally, such tools can be used in order to assess and improve the quality of medical services, as well as to compare risk profiles amongst distinct populations [16,17]. Although risk prediction models are not specifically designed to calculate the risk of complications for individual patients, they have been largely used to help physicians on decision making, especially in the context of cardiothoracic surgery [10]. For all these possible applications, it is advisable to take into account the differences in the population features before using the models at sites far from that of the original cohort. It is acknowledged that differences in baseline variables, ethic (or more specifically, genetically determined) characteristics, as well as environmental influences, might result in significant diversions in comparison to the original source cohort. As an example, Yap and colleagues assessed the use of the EuroSCORE in an Australian cohort different from the derivation cohort, and the calibration of the model in these new patients was considered poor [18].

Ideally, scores for risk estimation should be developed at each specific location, considering all the local genetic and environmental issues. Previous studies have investigated locally developed risk indexes in Brazilian populations, with variable success rates that are worth to be mentioned. Almeida et al. [19] have investigated distinct parameters as determinants of elevated risk after isolated CABG, but no prediction rule was proposed. Guarnaga et al. [20] proposed a risk score for patients submitted to surgery for valve replacement and the model has shown to be very sensitive to detect the risk of death. Gomes et al. [21] also proposed in 2007 a score system based on information collected before surgery and in the first postoperative day, coming up with eight variables that have shown useful to predict mortality. However, none of these prediction models have been disseminated to use in clinical practice so far.

The alternative approach might be to define the appropriateness of the scoring system of a previously developed model as described in this article. This means more specifically, to verify whether it can adequately fit to the clinical, demographic and environmental reality of the local population where it is intended to be used. Previously, a comparable approach was applied to assess the EuroSCORE in Brazilian population, showing satisfactory capability to predict death after CABG surgeries [22]. As we could observe after applying a similar methodology in our study the ROC curve has shown a highly significant area under the curve that represents a good discrimination for risk prediction for mortality and most of the studied endpoints.

Additionally, the goodness of fit test has indicated that progressively increasing levels of risk as estimated by STS-score was associated with comparable and also increasing mortality rates, reinforcing the appropriateness of the method in terms of calibration for the studied population. Similar results could be as well observed for general morbidity which represents a summary of all nine adverse outcomes. For the majority of the individual endpoints we could also observe
good performance for the calculated scores, with large and statistically significant areas under the ROC curves.

**Study limitations**

A few limitations of this study must be declared. First of all, our sample is not representative of the whole Brazilian population, as expected for a unicentric study. Second, the number of events for each endpoint is too small to unequivocally indicate the precision of the methods in our population. Additionally, for the endpoints stroke and mediastinitis, we were not able to calculate the accuracy using ROC curves, due to a low number of accounted events.

Furthermore, our analysis has shown a low capability to predict short length of hospital stay that is likely to be explained by the current use in our institution to discharge patients in the sixth or seventh post-operative day, meaning that almost none of the patients left the hospital prior to that time point. Presently, however, in view of the current international trends and the understanding that such practice could possibly impact short and long term morbidity, efforts have been made in order to shorten the hospital length of stay to below six days in our institution. This single action represents well how the use of the STS scoring system might impact clinical practice in terms of postoperative care, and additionally, contribute to reduction of complications.

**CONCLUSION**

Concluding, we have shown that STS scoring system is well calibrated to be used in the studied population that was submitted to CABG procedures, being able detect mortality and the majority of the investigated outcomes. In addition, the statistical methods used in our analysis cannot substitute an accurate and long term observation of our own population, using appropriate databanks to develop mathematical methods that will represent local genetic and environmental characteristics.

**REFERENCES**

1. Greason KL, Schaff HV. Myocardial revascularization by coronary arterial bypass graft: past, present, and future. Curr Probl Cardiol. 2011;36(9):325-68.

2. 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.

3. Anderson RP. First publications from the Society of Thoracic Surgeons National Database. Ann Thorac Surg. 1994;57(1):6-7.

4. Clark RE. The STS Cardiac Surgery National Database: an update. Ann Thorac Surg. 1995;59(6):1376-80.

5. O’Brien SM, Shahian DM, Filardo G, Ferraris VA, Haan CK, Rich JB, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2- isolated valve surgery. Ann Thorac Surg. 2009;88(1 Suppl):S23-42.

6. Shahian DM, O’Brien SM, Filardo G, Ferraris VA, Haan CK, Rich JB, et al. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 1--coronary artery bypass grafting surgery. Ann Thorac Surg. 2009;88(1 Suppl):S2-22.

7. http://209.220.160.181/STSWebRiskCalc261/. STS Calculator. 2011.

8. Chambless LE, Diao G. Estimation of time-dependent area under the ROC curve for long-term risk prediction. Stat Med. 2006;25(20):3474-86.

9. Anderson RJ, O’Brien M, MaWhinney S, VillaNueva CB, Moritz TE, Sethi GK, et al. Renal failure predisposes patients to adverse outcome after coronary artery bypass surgery. VA Cooperative Study #5. Kidney Int. 1999;55(3):1057-62.

10. Ferguson TB Jr, Hammill BG, Peterson ED, DeLong ER, Grover FL; STS National Database Committee. A decade of change: risk profiles and outcomes for isolated coronary artery bypass grafting procedures, 1990-1999: a report from the STS National Database Committee and the Duke Clinical Research Institute. Society of Thoracic Surgeons. Ann Thorac Surg. 2002;73(2):480-9.

11. Grover FL, Shroyer AL, Hammermeister KE, Edwards FH, Ferguson TB Jr, Dziuban SW Jr, et al. A decade’s experience with quality improvement in cardiac surgery using the Veterans Affairs and Society of Thoracic Surgeons national databases. Ann Surg. 2001;234(4):464-72.

12. Farsky PS, Graner H, Duccini P, Zandonadi EC, Amato VL, Anger J, et al. Risk factors for sternal wound infections and application of the STS score in coronary artery bypass graft surgery. Rev Bras Cir Cardiovasc. 2011;26(4):624-9.

13. Strabelli TM, Stolf NA, Uip DE. Practical use of a risk assessment
model for complications after cardiac surgery. Arq Bras Cardiol. 2008;91(5):342-7.

14. Ad N, Barnett SD, Speir AM. The performance of the EuroSCORE and the Society of Thoracic Surgeons mortality risk score: the gender factor. Interact Cardiovasc Thorac Surg. 2007;6(2):192-5.

15. Tzoulaki I, Liberopoulos G, Ioannidis JP. Assessment of claims of improved prediction beyond the Framingham risk score. JAMA. 2009;302(21):2345-52.

16. van Dieren S, Beulens JW, Kengne AP, Peelen LM, Rutten GE, Woodward M, et al. Prediction models for the risk of cardiovascular disease in patients with type 2 diabetes: a systematic review. Heart. 2012;98(5):360-9.

17. Braile DM, Monteiro R, Brandau R, Jatene FB. Risk prediction models: are they really necessary? Arq Bras Cardiol. 2010;95(6):677-8.

18. Yap CH, Reid C, Yii M, Rowland MA, Mohajeri M, Skillington PD, et al. Validation of the EuroSCORE model in Australia. Eur J Cardiothorac Surg. 2006;29(4):441-6.

19. Almeida FF, Barreto SM, Couto BR, Starling CE. Predictive factors of in-hospital mortality and of severe perioperative complications in myocardial revascularization surgery. Arq Bras Cardiol. 2003;80(1):51-60, 41-50.

20. Guaragna JC, Bodanese LC, Bueno FL, Goldani MA. Proposed preoperative risk score for patients candidate to cardiac valve surgery. Arq Bras Cardiol. 2010;94(4):541-8.

21. Gomes RV, Tura B, Mendonça Filho HT, Almeida Campos LA, Rouge A, Matos Nogueira PM, et al. A first postoperative day predictive score of mortality for cardiac surgery. Ann Thorac Cardiovasc Surg. 2007;13(3):159-64.

22. Andrade IN, Moraes Neto FR, Oliveira JP, Silva IT, Andrade TG, Moraes CR. Assessment of the EuroSCORE as a predictor for mortality in valve cardiac surgery at the Heart Institute of Pernambuco. Rev Bras Cir Cardiovase. 2010;25(1):11-8.