

UNIVERSITY OF WESTMINSTER



WestminsterResearch

<http://www.wmin.ac.uk/westminsterresearch>

Risk stratification in assessing risk in coronary artery bypass surgery.

Mike Rees¹

Jitesh Dineschandra²

¹ School of Informatics, University of Westminster

² Dendrite Clinical Systems, London

Copyright © [2006] IEEE. Reprinted from the proceedings of the 19th IEEE International Symposium on Computer-Based Medical Systems, IEEE CBMS 2006, Salt Lake City, USA, 22-23 June 2006. IEEE, Los Alamitos, USA, pp. 303-308. ISBN 0769525171.

This material is posted here with permission of the IEEE. Such permission of the IEEE does not in any way imply IEEE endorsement of any of the University of Westminster's products or services. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. By choosing to view this document, you agree to all provisions of the copyright laws protecting it.

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners. Users are permitted to download and/or print one copy for non-commercial private study or research. Further distribution and any use of material from within this archive for profit-making enterprises or for commercial gain is strictly forbidden.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of the University of Westminster Eprints (<http://www.wmin.ac.uk/westminsterresearch>).

In case of abuse or copyright appearing without permission e-mail wattsn@wmin.ac.uk.

Risk Stratification in Assessing Risk in Coronary Artery Bypass Surgery

Mike Rees¹, Jitesh Dineschandra²
¹Health and Social Care Modelling Group
Cavendish School of Computer Science
University of Westminster, London, UK
reesm@wmin.ac.uk

²Dendrite Clinical Systems, London, UK
Jitesh.Dineschandra@e-dendrite.com

Abstract

We present the need for risk stratification in the monitoring of cardiac surgical practice and review the frequentist and Bayesian approaches to the problem. Developments in the available databases are described. Enhancements to the Parsonnet and EuroSCORE systems are reviewed. We argue that in the UK, although the use of the Parsonnet system is inappropriate and that the EuroSCORE system is a clear improvement, there are advantages in adopting a system based on a Bayesian model for risk assessment.

1. Introduction

Techniques to facilitate the monitoring of cardiac surgical practice have been developing apace since data were first collected in 1977. Whilst performing statistical analysis on crude performance data has its uses it has long been realised that any form of comparative measures of performance would require the incorporation of risk adjustment to ensure that individual differences are fully accounted for. Unfortunately there is no agreement on exactly how that risk adjustment is to be performed.

Our main point of study, better prediction of patient outcomes, is influenced by many factors such as severity of illness, place of surgery and effectiveness of treatment. Studies must account for variability of these risk factors and in some cases it is possible to reduce or eliminate outcome variability through randomisation [1] which should balance known and unknown risk factors. Since most studies in medicine are observational rather than randomised, risk adjustment has been used to eliminate the effects of variability of

known risk factors on outcome measures. Statistical modeling techniques such as multivariate regression analysis can be used to study the association of various risk factors affecting a particular medical procedure and predict the impact of each risk factor on outcome [1]. Collectively these risk factors will then provide a predictive outcome measurement of patient risk.

We will concentrate solely on Coronary Artery Bypass Graft (CABG) procedures and draw on data from the National Adult Cardiac Surgical Database. We begin by sketching a background on the collection of such data.

We then review the initial approaches to risk adjustment resulting in the Parsonnet and EuroSCORE scoring systems and discuss recent revisions of these systems and assess their current importance for risk adjustment in the field of cardiac surgery.

As an alternative to the traditional, frequentist, approach alternative Bayesian models have been proposed [2], [3]. We describe the two main models, the 'simple' (5-factor) model and the 'complex' (9-factor) model, but concentrate our work on the use of the complex Bayes model.

Given the plethora of competing models there is a need to compare their performance. In this paper we produce a comparison of the various models using the Receiver Operating Characteristic (ROC) curve [4].

A number of studies have sought to compare various methodologies for risk assessment. We discuss our results in the context of other comparative reviews [5], [6], [7], [8], [9].

2. The SCTS databases

The Society of Cardio-Thoracic surgeons of Great Britain and Ireland (SCTS) have been capturing data for over 25 years. In 1977 The Society of Cardiothoracic Surgeons established the United Kingdom Cardiac Surgical Register (UKCSR) in order to collect activity and mortality data on all cardiac surgical procedures [10]. A report containing aggregated data was published every year and in 1997 individual surgeon's results were included for the first time. This activity was discontinued in 2003 to focus on the National Adult Cardiac Database; the 2003 report [11] was the first published report containing individual surgeon performance indicators.

In 1986 the Department of Health funded the establishment of the United Kingdom Heart Valve Registry. The success and value of this long-term initiative was instrumental in encouraging the Department of Health to proceed with the Central Cardiac Audit Database (CCAD) project, which aims to provide long-term follow up for all interventional cardiac procedures. The CCAD will provide an internationally unique opportunity for understanding the impact of both surgical and non-surgical interventions for acquired heart disease. Its strength lies in its linkage with the Office of National Statistics for long term mortality tracking which will aid the understanding of who will benefit most from which intervention.

3. The Parsonnet and EuroSCORE scoring systems

A standard approach to the problem of identifying risk factors would involve a multivariate statistical analysis on demographic patient records to identify those factors affecting the risk of the patient. This would typically involve a logistical regression analysis. It was this technique that resulted in the first validated method for adjusting mortality rates in 1985, the Parsonnet scoring system [12], [13].

The Parsonnet scoring system identified 16 risk factors as potentially having a bearing on the risk to a patient about to undergo cardiac surgery. These factors included the person's age, gender, whether they had a history of hypertension or diabetes and whether, for example, they were dialysis dependent. Once the patient data had been assembled scores were then mapped to the factors on the basis of the results, for example, a patient aged 73 years old would have their Parsonnet score increased by 7 whereas an 82 year old patient would find their score increased by 20. A

simple additive score from the results for all 16 factors is then calculated.

A weakness of the Parsonnet system is that two of the sixteen factors permit an element of subjectivity to enter into the scoring process, for example, a score in the range of 2 to 10 can be given to a patient presenting 'rare circumstances' such as severe asthma or paraplegia. In addition, whilst the weightings associated with each factor remained the same the impact of those factors was found to change with time. Consequently it was subsequently found that although the Parsonnet scoring system correlated well with a patient's risk the presence of these subjective factors coupled with these changes identified over time meant that the system tended to over predict a patient's risk. In the analysis that follows the subjective factors are removed from the model.

The Parsonnet system was developed in North America and was constructed using data gathered in that country. A similar methodology, but drawing on information based on a sample of pan-European cardiac surgery patients, lead to the creation of the EuroSCORE scoring system [14], [15]. This system specified 17 risk factors, none of which were subjective. A further variant is the logistic EuroSCORE model where non-integer weightings are given to each of the risk factors [11].

There have been many proposed variations to these systems with authors proposing the addition and/or removal of various factors from the system. A modification of the Parsonnet scoring system has been proposed for use in France which contained a total of 44 risk factors [16]. A study in the north west of England has argued for a reduction in the number of Parsonnet factors to nine [9]. Both studies were critical of the use of subjective factors and the inclusion of factors not associated with mortality. An improvement to the EuroSCORE model on replacing the serum creatine factor by creatine clearance has been reported [17].

4. Bayesian models

The proposed Bayesian models incorporate risk factors that introduce evidence into the model via conditional probabilities [2], [3].

Risk stratification is a method of delimiting sub-populations within a cohort that have different risks of a particular outcome, based upon severity of illness and morbidity. Using such an approach, it is possible to make fair comparisons between different institutions or different surgeons. Comparisons between individual institutions' or individual institutions' outcome rates

could also be made against agreed standards using this method [11].

One major recent trend is that of evidence-based medicine [18]. Evidence-based medicine requires an integrated assessment of the available evidence, and associated uncertainty, but there is also an emphasis on decision-making, for individual patients, or at other points in the health-care system. This demands consideration of the values and costs associated with potential outcomes.

4.1. The 9-factor Bayesian risk model

In the 1999-2000 National Adult Cardiac report [2] two Bayesian models were described to predict in-hospital mortality from the data on isolated coronary artery bypass surgery: a simple, 5-factor Bayes model and a complex, 9-factor Bayes model. Here we only discuss the 9-factor model, as it is equally applicable in the 5-factor model. The nine factors contained in this model are: Age, Body Surface Area, Diabetes, Hypertension, Left Main Stem disease, Ejection Fraction, Priority, Renal system and Previous operations.

Simple risk models can produce acceptable results, but more complex risk models may give better results. These two risk models were generated in an attempt to test this hypothesis. It has been demonstrated that whilst both models discriminated well, the complex model consistently showed slightly better discrimination than the simple model [11].

The aim was to generate risk models that accurately discriminated between patients who died following surgery and patients who survived; this was tested using Receiver Operating Characteristic (ROC) curve analysis [2][4][11]. Each Bayes model was initially trained [19] on the isolated coronary artery bypass surgery data from the financial year ending 1999, and then tested on the isolated coronary artery bypass surgery data from the financial years ending 1999 and 2000 separately.

4.2. Bayes tables

The Bayes table approach is a particularly simple way of building a risk stratification system from a database. Based solely on tables relating outcomes to single risk factors, the probability of an adverse outcome can be estimated for a patient with any combination of risk factors.

The method is based on the repeated use of Bayes theorem, which tells us how the probability of an event should be revised when additional relevant evidence is obtained.

Suppose that S represents survival, D represents death, A represents patients in the age grouping 70-79 years old and not A represents all other patients. A probability of an event is calculated as the number of events in a group (post-operative deaths) divided by the total number in that same group such that

$$P(D) = \frac{D}{S + D}$$

Furthermore, since survival and death must occur we have from the general laws of probability that

$$P(D) + P(S) = 1$$

The probability of survival can be calculated as $P(S) = 1 - P(D)$. The *odds* on death are defined as the probability of death divided by the probability of survival and can be denoted by

$$\begin{aligned} odds(D) &= \frac{P(D)}{P(S)} \\ &= \frac{P(D)}{[1 - P(D)]} \end{aligned}$$

The odds on D , equation (2.6) may be converted to a probability using the following formula

$$P(D) = \frac{odds(D)}{[1 + odds(D)]} \quad (4.1)$$

Suppose that we now wish to take into account the age of the patient, which falls in the range 70-79 years old (designated A). The probability of death in this group is denoted by $P(D | A)$ and this is known as the posterior probability. The odds on death in this group are $odds(D) = P(D | A) / P(S | A)$ and this is known as the posterior odds.

Bayes theorem is the formula that provides the link between the prior and posterior odds, that is

$$posterior\ odds = prior\ odds \times likelihood\ ratio \quad (4.2)$$

where the likelihood ratio expresses how much more likely it is that a patient with such an age should fall amongst those who die rather than those who survive.

Items of evidence in a Bayes model might include many different risk factors that affect the outcome, such as age, left ventricular ejection fraction, urgency

and so on. The calculation of risk for a specific patient is updated each time a new item of evidence is added. This simple procedure is, however, making the crucial assumption that each item of evidence is contributing independent information concerning the chances of the outcome; technically, we are assuming that the items of evidence are conditionally independent given the true outcome. This assumption is most likely to be appropriate if careful clinical sense has been used in selecting the predictive factors that do not convey similar evidence. Failure to select predictive factors in this way may result in predictions that are over-confident.

It is important to note that the score is calculated irrespective of omissions in the data. It is possible to calculate a Bayes score whether all, some or none of the risk factors are known. The Bayes score is adjusted for each item of evidence, and will more accurately reflect the true risk if all the relevant data are known.

4.3. Calculating individual patient Bayes scores

Suppose we have a 66-70 year old patient undergoing isolated coronary artery bypass surgery. Referencing the Bayes table for the complex 9-factor Bayes model in [11], the prior odds are 0.026 and the likelihood ratio for a 66-70 year old patient is 1.09. The posterior odds of death for this patient can be calculated using equation (4.2) as follows

$$\text{Posterior odds on death} = 0.026 \times 1.09 = 0.028$$

Using equation (4.1) the posterior probability of death for this patient is

$$[0.028 / (1 + 0.028)] = 0.027$$

or 2.7%.

This can be extended to use all of the nine factors as described in section 2. If this patient was diabetic, and again using the Bayes table for the complex 9-factor Bayes model in [11], the likelihood ratio for a diabetic patient is 1.29. Using equations (4.1) and (4.2) the posterior odds on death is $0.026 \times 1.09 \times 1.29 = 0.037$ and the posterior probability of death is 0.036 giving a posterior percentage risk of death of 3.6%.

It is also possible to extend this model to incorporate a damping factor in the posterior odds calculation, which gives a more conservative percentage risk of death [2].

5. Use of the receiver operating characteristic curve

The area under an ROC curve represents the probability that the risk predictor accurately discriminates between patients who die during surgery and patients who survive surgery [4]. An area of 0.50 indicates that there is no discrimination. An area of 1.00 would indicate that discrimination was perfect. The closer the value is to 0.50 the less accurate the discrimination, and the closer to 1.00 the better the discrimination.

It is also important that risk-scoring models cannot only estimate risk for individual patients but also for a group of patients. Calibration plots can be used to test whether the risk model estimates risk for both individual patients as well as for groups of patients. To generate a calibration plot we may plot the observed number of events against the predicted number of events by the model. When plotted side by side, the closer the match between both, the more accurate is the risk model.

Data quality is crucial in order to compare such risk-scoring models, taking into account the limitations of each model being compared, as this will increase the validity of each of the models being compared. Figure 1 shows a calibration plot of Chi-square values for four different risk models: Parsonnet, Logistic EuroSCORE, EuroSCORE and Complex Bayes score.

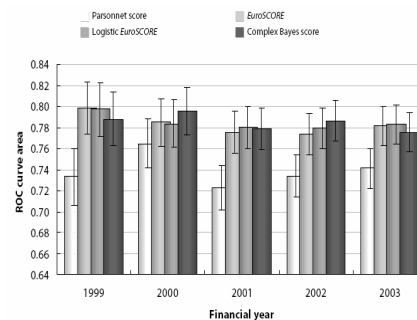


Figure 1. Isolated CABG: Calibration plot chi-squared values for 4 different risk models (n=43,822)

The chi-squared value is a measure of how closely the predicted mortality matches the actual mortality. The smaller the chi-squared value, the better the match.

The key messages are that the match was good for both the EuroSCORE and the Complex Bayes score, but, as time goes by, the match between observed and predicted mortality slowly drifts apart for all the scores. Figure 2 demonstrates an ROC analysis [2] [4] [11] of the four risk-scoring models mentioned above.

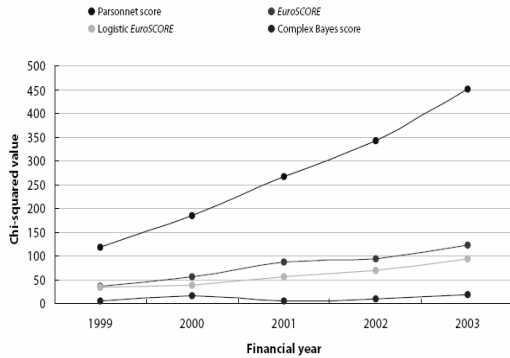


Figure 2. Isolated CABG: ROC curve areas for 4 different risk-models (n=43,819)

Figure 2 indicate that the accuracy of all risk prediction models varies with time, but that the Complex Bayes Score is consistently the most accurate model for prediction of death following coronary surgery in the UK.

6. A comparative review

The Parsonnet scoring system performs well when applied in North America and has been shown to outperform other general severity systems.

Attempts to modify the Parsonnet scoring system for use in France and England have resulted in some improvement; increasing the area under the ROC curve from 0.65 to 0.71 and from 0.68 to 0.73 respectively [9], [16]. However, such improvements still leave the Parsonnet system being outperformed by the EuroSCORE and Bayes systems [11].

Other American systems also have a poorer performance record when it has been attempted to use them in a UK setting [6].

7. Conclusions and future work

Attempts to modify the North American Parsonnet system for use in the UK have not resulted in sufficient improvement to advocate its use in preference to the EuroSCORE and Bayes systems.

Although there seems little to choose between the EuroSCORE and Bayes systems in terms of discrimination as measured by the area under the ROC curve, calibration plots suggest the Bayes system to be more stable over time.

In addition to the use of Bayes tables a clearer and potentially more accessible approach could be to model these processes through the use of Bayesian networks. Further research into the use of Bayesian classification

algorithms may result in a greater explanatory power of the Bayes systems.

8. References

- [1] D.M Shahian, E.H. Blackstone et. al. "Cardiac surgery risk models: a position article," *Annals of Thoracic Surgery*, 2004, 78(5), pp. 1868 – 1877
- [2] B.E. Keogh, and R. Kinsman, *National Adult Cardiac Surgical Database Report 1999 - 2001*, Dendrite Clinical Systems, 2003.
- [3] D. Majumdar, "An axiomatic characterization of Bayes' rule," *Mathematical Social Sciences*, 2004, 47(3), pp. 261 - 273.
- [4] J.A. Hanley and B.J. McNeil, "The meaning and use of the area under a Receiver Operating Characteristic curve," *Radiology*, 1982, 143, pp. 29 - 36.
- [5] G. Asimakopoulos, S. Al-Ruzzeh, et.al., "An evaluation of existing risk stratification models as a tool for comparison of surgical performances for coronary artery bypass grafting between institutions," *European Journal of Cardio-Thoracic Surgery*, 2003, 23 (6), pp. 935 - 942.
- [6] B. Bridgewater, H. Neve, et. al., "Predicting operative risk for coronary artery surgery in the United Kingdom: a comparison of various risk prediction algorithms," *British Heart Journal*, 1998, 79(4), pp. 350 - 355.
- [7] J. Martinez-Alario, I.D. Tuesta, et. al., "Mortality prediction in cardiac surgery patients: comparative performance of Parsonnet and general severity systems," *Circulation*, 1999, 99(18), pp. 2378 - 2382.
- [8] P.C. Austin, C.D. Naylor, et.al., "A comparison of Bayesian vs. a frequentist method for profiling hospital performance," *Journal of Evaluation in Clinical Practice*, 2001, 7(1), pp. 35 - 45.
- [9] K. Wynne-Jones, M. Jackson, et. al. "Limitations of the Parsonnet score for measuring risk stratified mortality in the north west of England," 2000, *Heart*, 84(1), pp. 71 – 78.
- [10] T.A.H. English, A.R. Bailey et. al. "The UK cardiac surgical register 1977-82," *British Medical Journal*, 1984, 289, pp. 1205 – 1208.
- [11] B.E. Keogh, and R. Kinsman, *National Adult Cardiac Surgical Database Report 2003: improving outcomes for patients*, Dendrite Clinical Systems, 2003.
- [12] V. Parsonnet, D. Dean and A.D. Bernstein, "A method for uniform stratification of risk for evaluating the results of

surgery in acquired heart disease,” *Circulation*, 1989, 79, pp. 13 - 112.

[13] A.D. Bernstein and V. Parsonnet, “Bedside estimation of risks as an aid for decision-making in cardiac surgery,” *Annals of Thoracic Surgery*, 2000, 69(3), pp. 823 - 828.

[14] S.A.M. Nashif, F. Roques et. al. and the EuroSCORE Study Group European system for cardiac operative risk evaluation (EuroSCORE), presented to the 12th Annual Meeting of the European Association for Cardio-Thoracic Surgery, 1998.

[15] F. Roques, S.A.M. Nashef, et. al. “Risk factors and outcome in European cardiac surgery: analysis of the EuroSCORE multinational database of 19030 patients,” *European Journal of Cardio-thoracic Surgery*, 1999, 15(6), pp. 816 – 823.

[16] F. Gabrielle, F. Roques, F., et. al. (1997) “Is the Parsonnet's score a good predictive score of mortality in adult cardiac surgery: assessment by a French multicentre study,” *European Journal of Cardio-Thoracic Surgery*, 1997, 11 (3), pp. 406 - 414.

[17] J. Walter, A. Mortasawi, et. al., “Creatine clearance versus serum creatine as a risk factor in cardiac surgery,” 2003, *BMC Surgery*, 3(1); p. 4.

[18] A.F.M Smith, D. Ashby, “Evidence-based medicine as Bayesian decision-making,” *Statistics in Medicine*, 2000, 19(23), pp. 3291 – 3305.

[19] D.G. Seymour, M. Green, F.G. Vaz, “Making better decisions: construction of clinical scoring systems by the Spiegelhalter-Knill-Jones approach,” *British Medical Journal*, 1990, 300(6719), pp. 223 – 226.