Prediction of in-hospital death following aortic valve replacement: a new accurate model

Matthew Richardson*, Neil Howellb,c, Nick Freemantled, Ben Bridgewaterf and Domenico Pagano* on behalf of The Quality and Outcomes Research Unit University Hospital Birmingham

Received 30 December 2011; received in revised form 27 June 2012; accepted 2 July 2012

Abstract

OBJECTIVES: Aortic valve replacement (AVR) is accepted as the standard treatment for severe symptomatic aortic valve stenosis and regurgitation. As novel treatments are introduced for patients at high risk for conventional surgery, it is important to have models that accurately predict procedural risk. The aim of this study was to develop and validate a risk-stratification model to predict in-hospital risk of death for patients undergoing AVR and to compare the model with existing algorithms.

METHODS: We reviewed data from the Central Cardiac Adult Database, which holds prospectively collected clinical information on all adult patients undergoing cardiac surgery in National Health Service (NHS) hospitals and some private providers in the UK and Ireland. We included all the patients undergoing AVR with or without coronary artery bypass grafting. The study population consists of 55 157 patients undergoing surgery between 1 April 2001 and 31 March 2009. The model was built using data from April 2001 to March 2008 and validated using data from patients undergoing surgery from April 2008 to March 2009. The model was compared against the additive and logistic EuroSCORE models and a valve-specific risk-prediction model.

RESULTS: The final multivariable model includes items describing cardiovascular risk status and procedural factors. Applying the model to the independent validation dataset provided a c-statistic (index of rank correlation) of 0.791, which was substantially better than that achieved by previously developed risk models in Europe, and significantly improved risk prediction in higher-risk patients.

CONCLUSIONS: We have produced an accurate risk model to predict outcome following AVR surgery. It will be of use for patient selection and informed consent, and of particular interest in defining those patients at high risk who may benefit from novel approaches to AVR.

Keywords: Risk model · Risk score · Prognostic model · Aortic valve replacement

INTRODUCTION

Aortic valve replacement (AVR) is an established treatment for aortic valve disease and is a Class I indication for patients with symptomatic severe aortic stenosis or regurgitation [1]. Successful surgery improves both symptoms and life expectancy in these patients. In a recent report based on 30 000 patients undergoing cardiac surgery between 2004 and 2008 in the UK, the overall in-hospital mortality for isolated AVR was 2.8%, increasing from 1.7% in patients <65 years of age up to 5.8% in those aged between 81 and 85 years [2]. Providing informed consent for patients requires the accurate prediction of procedural risk, and accurate risk-prediction models are also important for comparing outcomes between hospitals or surgeons to support quality assurance and quality improvement processes [3]. Recent reports have suggested that the ability of existing models to predict risk is declining. [2, 4–7]. Traditional risk-predicting models overestimate operative mortality in the contemporary patient population, and this is particularly evident in the high-risk group. A recent report from the European Society of Cardiology highlighted the specific problems faced in the accurate outcome prediction of elderly high-risk patients, in whom trans-catheter aortic valve implantation (TAVI) [8–11] may be considered as an alternative to prohibitive-risk conventional surgery. The poor performance of current prediction models [12, 13] in this context is therefore of great importance, as tailoring the appropriate strategy to the individual patient remains challenging. Furthermore, the lack of an accurate prediction...
algorithm makes it difficult to compare outcomes between conventional surgery and novel techniques. We have therefore developed a new model for estimating in-hospital mortality after conventional aortic valve surgery using the large National Surgical Register of the Society for Cardiothoracic Surgery in GB and Ireland and compared it to existing risk-prediction algorithms that use similar variables.

METHODS

Data source

All patients in the Society for Cardiothoracic Surgery in GB and Ireland adult cardiac surgery database undergoing conventional AVR with or without coronary artery surgery from April 2001 to March 2008. Data on 55,157 patients were used as the development sample to create a model for predicting the risk of in-hospital mortality. Data on 13,277 patients undergoing surgery between April 2008 and March 2009 were then used as the validation sample. The Society for Cardiothoracic Surgery in GB and Ireland (SCTS) minimum dataset was collected by all hospitals and returned to the Central Cardiac Audit Database [CCAD (www.ccad.org.uk)], a part of the NHS Information Centre (www.nhsic.nhs.uk). In-hospital mortality is included in this dataset.

Statistical analysis and model validation

A set of prespecified candidate variables selected by the study group on the basis of prior knowledge and clinical experience were used to develop models for in-hospital mortality. Missing data were accounted for through multiple imputations, which were performed using the IVEware software (http://www.isr.umich.edu/src/smp/ive/). In-hospital survival was modelled using multivariable non-linear mixed models with a logit link and Binomial/Gaussian error, the dependent variable being the binary outcome death or discharge. The effect of the surgical team was addressed through fitting a random effect on this stratum. Forward stepwise selection was used to obtain a subset of statistically significant ($\alpha = 0.05$) variables. A model with these variables as main effects plus all first order interactions was then produced. Further forward, stepwise selection was applied to reduce the dimension of this model. The following statistical software routines were used: PROC LOGISTIC and PROC GLIMMIX in SAS v9.2 (http://www.sas.com) and the Design package in GNU R v 2.9.2 (http://www.R-project.org). The model for in-hospital mortality was validated on the development and the validation samples using the c-statistic. The Hosmer–Lemeshow test was used in order to assess model heterogeneity by partitioning the cohort into 10 risk strata of similar size and comparing the difference between the observed and expected number of events. A more detailed description of the statistical methodology, analysis of missing data and the model coefficients are available in the online supplement.

Comparison with existing models

The performances of the additive EuroSCORE, logistic EuroSCORE and the score developed by Ambler 2005 [14, 15] were compared against the current model for in-hospital mortality. Logistic regression with the above scores as the independent variable was performed on the dataset for April 2008 to March 2009. Head-to-head direct comparison with the Society of Thoracic Surgeons (STS) score was not possible as this uses different variables, which are not routinely collected by the SCTS minimum dataset.

RESULTS

Descriptive statistics are shown in Supplementary Tables A1 and A2. Mortality for patients undergoing isolated AVR was 3.4% (990/28,981), and for those undergoing AVR + coronary artery bypass grafting (CABG) it was 6.1% (1182/19,222). Mortality in the cohort as a whole was 5.2% (2864/54,975). Missing data are accounted for through multiple imputations, which were performed using the IVEware software (http://www.isr.umich.edu/src/smp/ive/). In-hospital survival was modelled using multivariable non-linear mixed models with a logit link and Binomial/Gaussian error, the dependent variable being the binary outcome death or discharge. The effect of the surgical team was addressed through fitting a random effect on this stratum. Forward stepwise selection was used to obtain a subset of statistically significant ($\alpha = 0.05$) variables. A model with these variables as main effects plus all first order interactions was then produced. Further forward, stepwise selection was applied to reduce the dimension of this model. The following statistical software routines were used: PROC LOGISTIC and PROC GLIMMIX in SAS v9.2 (http://www.sas.com) and the Design package in GNU R v 2.9.2 (http://www.R-project.org). The model for in-hospital mortality was validated on the development and the validation samples using the c-statistic. The Hosmer–Lemeshow test was used in order to assess model heterogeneity by partitioning the cohort into 10 risk strata of similar size and comparing the difference between the observed and expected number of events. A more detailed description of the statistical methodology, analysis of missing data and the model coefficients are available in the online supplement.

Risk calculator

A risk calculator was developed to predict risk for in-hospital mortality. The variables used in the risk calculator are shown in Table 3.

<table>
<thead>
<tr>
<th>Risk group</th>
<th>n</th>
<th>Observed deaths</th>
<th>Expected deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (3.72e-05, 0.00837)</td>
<td>1320</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>2 (0.00837, 0.0125)</td>
<td>1317</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>3 (0.0125, 0.0165)</td>
<td>1308</td>
<td>19</td>
<td>19</td>
</tr>
<tr>
<td>4 (0.0165, 0.021)</td>
<td>1306</td>
<td>21</td>
<td>25</td>
</tr>
<tr>
<td>5 (0.021, 0.0262)</td>
<td>1300</td>
<td>25</td>
<td>31</td>
</tr>
<tr>
<td>6 (0.0262, 0.0333)</td>
<td>1285</td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>7 (0.0333, 0.0434)</td>
<td>1262</td>
<td>62</td>
<td>50</td>
</tr>
<tr>
<td>8 (0.0434, 0.0602)</td>
<td>1267</td>
<td>59</td>
<td>68</td>
</tr>
<tr>
<td>9 (0.0602, 0.0949)</td>
<td>1210</td>
<td>110</td>
<td>99</td>
</tr>
<tr>
<td>10 (0.0949, 0.998)</td>
<td>1073</td>
<td>249</td>
<td>251</td>
</tr>
</tbody>
</table>

Risk groups are formed by splitting predicted risk of 30-day mortality into 10 equal groups, risk group 1 contains the low-risk patients (risk of 30-day mortality 0.0000372–0.00837), risk group 10 contains the high-risk patients (risk of 30-day mortality 0.0949–0.998). The observed number deaths is obtained by counting the number of actual deaths for patients within a particular risk group, the expected number deaths is obtained by summing the predicted risk of 30-day mortality for all patients within a particular risk group. (On performing the Hosmer–Lemeshow for the validation set we obtained, $X^2_{HL} = 8.1745, df = 8, P = 0.4162$).
A value close to 1 for the c-statistic indicates that the model has good discriminatory power.

Table 2: Performance of scores

<table>
<thead>
<tr>
<th>Score</th>
<th>c-statistic (index of rank correlation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New model</td>
<td>0.791</td>
</tr>
<tr>
<td>Logistic EuroSCORE</td>
<td>0.590</td>
</tr>
<tr>
<td>EuroSCORE</td>
<td>0.638</td>
</tr>
<tr>
<td>Ambler score</td>
<td>0.660</td>
</tr>
</tbody>
</table>

DISCUSSION

In this study, we have developed a clinically applicable model for the prediction of in-hospital mortality specifically for patients undergoing AVR, using the large contemporaneous national dataset from the UK. The model has a good discriminatory ability and is well calibrated, fitting the data well across all risk strata. The model is equally valid for patients undergoing surgery for stenosis or regurgitation, and for those who are undergoing concomitant procedures including revascularization. This study was based on a large national database of patients undergoing cardiac surgery. This database, which is used for a high-profile national audit project, has the confidence of clinicians and has a low incidence of missing data, and as such we believe it is fit for the purposes to which is has been applied [2].

Comparison with other models

Models widely used in risk stratification, the STS score, the EuroSCORE and the Ambler score all perform well as tools to rank patients according to risk, but the EuroSCORE is based on data from 1995 and is now poorly calibrated for assessing contemporary European data [16]. The STS score significantly out-performs the other models with a c-statistic of 0.799, a similar result for our current model, c-statistic = 0.791 [17]. There is, however, limited data on the homogeneity of these models and concern in particular has been raised over their predictive power in high-risk patients. The predictive power of the above models was examined by Dewey et al. [18] in a group of 638 patients identified as being at the highest risk decile strata. Each model identified a slightly different cohort at high risk. Observed vs predicted mortality was 13.8 vs 18.8% with the STS, 50.9 vs 15.6% for the logistic EuroSCORE, 14 vs 11.9% for the additive EuroSCORE and 19 vs 13.4% for the Ambler score, with logistic regression analysis demonstrating that the STS score had the highest odds ratio for predicting mortality in this high-risk group. During the design of our model, we paid specific attention to the problems of predicting mortality in high-risk patients, the issues of data that are important in predicting outcome but not routinely collected, and the problems of missing data by using multiple imputation for the missing data. Our model identified 1073 patients in the highest decile risk strata, predicting 249 deaths (23.2%) vs 251 observed deaths (23.4%). This gives an observed predicted mortality ratio of 1:1 for this score, 1:1.4 for the published STS score, and 0.85, 0.71 and 0.31 for the additive EuroSCORE, Ambler score and logistic EuroSCORE, respectively; our model displays a 20% improved prediction over logistic EuroSCORE, 16% improved prediction over the additive EuroSCORE and a 13% increased prediction over the model developed by Ambler et al. [15].

Interactions terms

We believe that the use of interaction terms during the statistical development of the model has been a major contributor in producing a highly predictive model. Unlike the previously published models referred to earlier, in addition to the usual clinical risk variables determined to be associated with mortality, we have also looked at first-order interaction terms, which provides a statistical solution to issues that are known to be clinically associated with increased risk, and as well as improving the statistical performance of the model, they also increase its face validity. For example, an elderly patient with poor ventricular function undergoing AVR will have a greater chance of survival than the patient with a poor ventricle whose pathology is a combination of ischaemia and end-stage hypertrophy. In our model, the interaction term Cardiac Procedure * Extent of Coronary Vessel Disease was significant, allowing us to vary the interaction term to provide a measure of the effect of undergoing isolated AVR or undergoing AVR + CABG varying degrees of coronary vessel disease. The inclusion of this interaction term Cardiac Procedure * Extent of Coronary improved the overall fit of the model.

Problems of prediction in high-risk patients

A significant problem encountered when predicting mortality in high-risk patients is that most patients in the high-risk strata have very similar risk profiles resulting in limited inter-patient variability upon which to discriminate. This problem is further complicated by the lack of data available on other potentially uncommon risk factors that are never-the-less associated with significant operative risk including malignancy and previous radiotherapy, liver cirrhosis, home oxygen therapy, frailty and malnutrition, and also variables with which considerable predictive power has been lost by their transformation from continuous data to medically convenient categorical data, e.g. creatinine >200 µmol l⁻¹ rather than serum creatinine or estimated glomerular filtration rate [19] that have been shown to be strong and independent predictors of operative risk. Finally, the model is
derived from patients who have actually undergone surgery and therefore its application in the assessment of risk in a sub-group of patients not previously referred for surgery may be limited. Such data are not available on the numbers of patients required to produce such a model, so to compensate for the lack of this data in developing the model, we sought to extrapolate beyond the data by using the surgical team as a random effect as a surrogate marker of the unobserved characteristics of patients and clinicians. This allows the variation in risk to be taken into account without having to assume some specific form for the relationship between risk and surgeon. This random effect accounts in part for some of the unobserved characteristics of patients and clinicians that the dataset does not account for. Although the surgeon cannot be included as a variable in the calculator, it influences the fixed effects that are a fundamental part of this model.

Missing data

Although the incidence of missing data is low, if the level of missing data for a particular variable is small, when considering the complete case analysis, the resulting loss of data can be substantial. A high level of missing data may lead to a model that is unreliable and misleading. We used multiple imputation to overcome the potential bias introduced by missing data and the observation that the model achieves a good discrimination on the independent validation data suggests that the technique has been successful.

Application for trans-catheter aortic valve implantation

This model has been specifically constructed for surgical AVR and therefore its application for the prediction of mortality in patients undergoing TAVI is not validated. It will, however, be clinically important in more accurately predicting surgical risk in the patient cohort in whom TAVI may be considered. It is, however, important to recognize the potential of this model to under-predict risk in patients with a conventionally high-risk profile who also present with other significant comorbidities not represented by the model. It may be important to collect accurate preoperative and outcome data on such patients to facilitate the construction of TAVI-specific risk models.

Conclusion

We have produced a contemporary procedure-specific model, based upon a limited number of familiar variables focusing on 30-day mortality. It has been validated on the contemporary results of aortic valve surgery in the UK and has been shown to be well-calibrated in all risk groups including the high-risk patients. As such, this model fulfils most of the criteria set by the European Society of Cardiology Working Group on Valvular Heart Disease Position Paper: assessing the risk of interventions in patients with valvular heart disease [20], and it may be of use in determining the optimum strategy for high-risk patients with aortic valve disease.

Unanswered question and future research

The model constructed in this study was based on data from UK patients, and it is unknown as to how it would perform on other international databases. While we can predict more accurately the outcomes of high-risk patients when considering conventional AVR, we do not know whether this model can be used to predict the outcome of patients undergoing TAVI procedures. Future refinement of this tool may more accurately inform decisions on the treatment of choice for aortic valve pathology and choice of prosthesis. Improvements in UK and European risk-prediction models may require the systematic collection of the data described above not currently collected.

SUPPLEMENTARY MATERIAL

Supplementary material is available at EJCTS online.

ACKNOWLEDGEMENTS

We would like to acknowledge all the surgical units, the surgeons and the database managers without whose tireless work the data to conduct this analysis would not have been possible. We would also like to acknowledge David Cunningham and his colleagues at the central cardiac audit database for their work in collating and extracting the data for this study.

Funding

This study was funded by the NHS East Midlands Specialised Commissioning Group.

REFERENCES


