Risk Prediction Models for Long-Term Survival after Cardiac Surgery: A Systematic Review

Leah Argus[1] Marcus Taylor[2] Maral Ouzounian[3] Rajamiyer Venkateswaran[2] Stuart W. Grant[4,5]

[1] The University of Manchester, Manchester, United Kingdom
[2] Department of Cardiothoracic Surgery, Manchester University NHS Foundation Trust, Manchester, United Kingdom
[3] Division of Cardiovascular Surgery, Peter Munk Cardiac Centre, University Health Network, Toronto, Ontario, Canada
[4] Division of Cardiovascular Sciences, University of Manchester, Manchester, United Kingdom
[5] Academic Cardiovascular Unit, South Tees Hospitals NHS Foundation Trust, Middlesborough, United Kingdom

Address for correspondence Leah Argus, MBChB (Hons), MRes (Dist), The University of Manchester, Manchester, United Kingdom (e-mail: leah.argus@nhs.net).

Thorac Cardiovasc Surg 2024;72:29–39.

Abstract

Background The reporting of alternative postoperative measures of quality after cardiac surgery is becoming increasingly important as in-hospital mortality rates continue to decline. This study aims to systematically review and assess risk models designed to predict long-term outcomes after cardiac surgery.

Methods The MEDLINE and Embase databases were searched for articles published between 1990 and 2020. Studies developing or validating risk prediction models for long-term outcomes after cardiac surgery were included. Data were extracted using checklists for critical appraisal and systematic review of prediction modeling studies.

Results Eleven studies were identified for inclusion in the review, of which nine studies described the development of long-term risk prediction models after cardiac surgery and two were external validation studies. A total of 70 predictors were included across the nine models. The most frequently used predictors were age ($n = 9$), peripheral vascular disease ($n = 8$), renal disease ($n = 8$), and pulmonary disease ($n = 8$). Despite all models demonstrating acceptable performance on internal validation, only two models underwent external validation, both of which performed poorly.

Conclusion Nine risk prediction models predicting long-term mortality after cardiac surgery have been identified in this review. Statistical issues with model development, limited inclusion of outcomes beyond 5 years of follow-up, and a lack of external validation studies means that none of the models identified can be recommended for use in contemporary cardiac surgery. Further work is needed either to successfully externally validate existing models or to develop new models. Newly developed models should aim to use standardized long-term specific reproducible outcome measures.

Keyword
- Cardiac
- outcomes
- surgery
- complications
- statistics

Introduction

In-hospital mortality rates after cardiac surgery in the United Kingdom (UK) have been published at the hospital and individual surgeon levels since 2005.[1] Since this time, despite an overall increase in risk profile, in-hospital mortality has continued to fall and has been as low as 0.6% following elective coronary artery bypass grafting (CABG) in recent years.[2] As in-hospital mortality declines and stabilizes,

received
June 23, 2022
accepted after revision
November 24, 2022
article published online
February 7, 2023

DOI https://doi.org/10.1055/s-0043-1760747.
ISSN 0171-6425.

measuring additional postoperative markers of quality such as postoperative length of stay, reoperation for bleeding, stroke, and renal failure is increasingly important.[3]

In addition to short-term procedural outcomes, long-term outcomes are an important quality marker as the majority of patients undergo cardiac surgery for late prognostic benefits in addition to symptomatic benefits. The increasing availability of less-invasive treatment options as an alternative to conventional cardiac surgery is another driver for shifting focus away from short-term procedural results to long-term outcomes.[4–6] If patients are going to make informed decisions about treatment options, then accurate information on long-term outcomes is imperative.

While informing treatment decisions is a major role of risk prediction models, they can also be used to undertake risk-adjusted outcome analyses. This enables fair comparisons of outcomes reducing the risk of inappropriate risk aversion. In the United Kingdom, published in-hospital mortality outcomes are risk adjusted using the EuroSCORE models, whereas in North America, the Society of Thoracic Surgeons scores are used for a similar purpose.[7–9]

Looking to the future, if clinical governance analyses of long-term cardiac surgery outcomes are to be considered, then similar risk-adjustment models would need to be available or developed. The objective of this study was to systematically review and assess risk prediction models developed specifically for the prediction of long-term outcomes after cardiac surgery to determine if any are potentially suitable for use in contemporary cardiac surgical practice.

Material and Methods

International Review Board approval or waiver, consent statement, and clinical trial registration were not applicable for this study.

Search Strategy and Data Sources

This systematic review was registered with PROSPERO (International Prospective Register of Systematic Reviews, CRD42021251462) and was undertaken in conjunction with a medical librarian. MEDLINE (searched using OVID, the online library of databases) and Embase were searched separately for articles between January 1990 and August 2020. The start date of 1990 was selected to ensure that only contemporary studies were included. The search terms specified were "cardiac surgery" AND "mortality" OR "survival," in addition to terms for risk prediction models as recommended by Geersing et al.[10] The complete search strategy is detailed in ►**Supplementary Table S1** (available online only).

Study Inclusion

All studies that described the development or validation of risk prediction models specifically designed to predict long-term outcomes after cardiac surgery were included. We defined "long term" as referring to all outcomes beyond 30-day and/or in-hospital mortality. Models measuring mortality, morbidity, quality of life, or any composite endpoint thereof were included. Studies evaluating noncardiac surgery, pediatric cardiac surgery, ventricular assist device implantation, cardiopulmonary transplantation, or diagnostic techniques were excluded. Models previously developed to predict short-term outcomes and, subsequently, used to predict long-term outcomes were excluded. Studies that examined the impact of individual predictors on outcomes but did not develop a model were also excluded. Only papers written in English were included.

Data Extraction and Quality Assessment

Manual screening of all articles identified in the initial search was performed by two independent reviewers (L.A. and M.T.). A further discussion with S.W.G. took place in the event of disagreement about the inclusion of specific articles. The titles, abstracts, and then full text were reviewed in sequential stages, to determine suitability for inclusion in the review (L.A. and M.T.). The reference lists of the eligible studies were searched to identify additional studies. A data extraction proforma was created based on the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies checklist and the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis Statement.[11,12] This includes information about participants, outcomes, candidate predictors, missing data, model development, model performance, model evaluation, results, and discussion and is summarized in ►**Supplementary Table S2** (available online only). Data quality was critically appraised based on the Methodological Index for Nonrandomized Studies (MINORS) instrument.[13] The risk of bias was analyzed using the Prediction Model Risk of Bias Assessment tool (PROBAST).[14]

Model Analysis

Model performance data extracted included both measures of discrimination and calibration. The measure of discrimination (the extent to which predicted risks discriminate between patients with and without the outcome) extracted was the area under the receiver operating characteristic curve (AUC), with an AUC value of >0.7 taken to represent acceptable discrimination. Measures of calibration (the extent to which predicted risk corresponds to observed risk) extracted included the Hosmer–Lemeshow (H–L) test, observed to expected (O:E) ratios and calibration plots. Scatterplots with linear regression and 95% confidence intervals were fitted for model AUCs against model end-point time.

Results

Selected Studies

The database search yielded 2,403 results. Of these, 11 studies met the inclusion criteria and were included in the review. Nine studies described the development of a model to predict long-term outcomes after cardiac surgery, while the other two were external validation studies. All studies were retrospective in nature and the median sample size was 9,393 (range, 2,031–348,341). The study selection process is detailed in ►**Fig. 1** and model performance is summarized in ►**Table 1**.

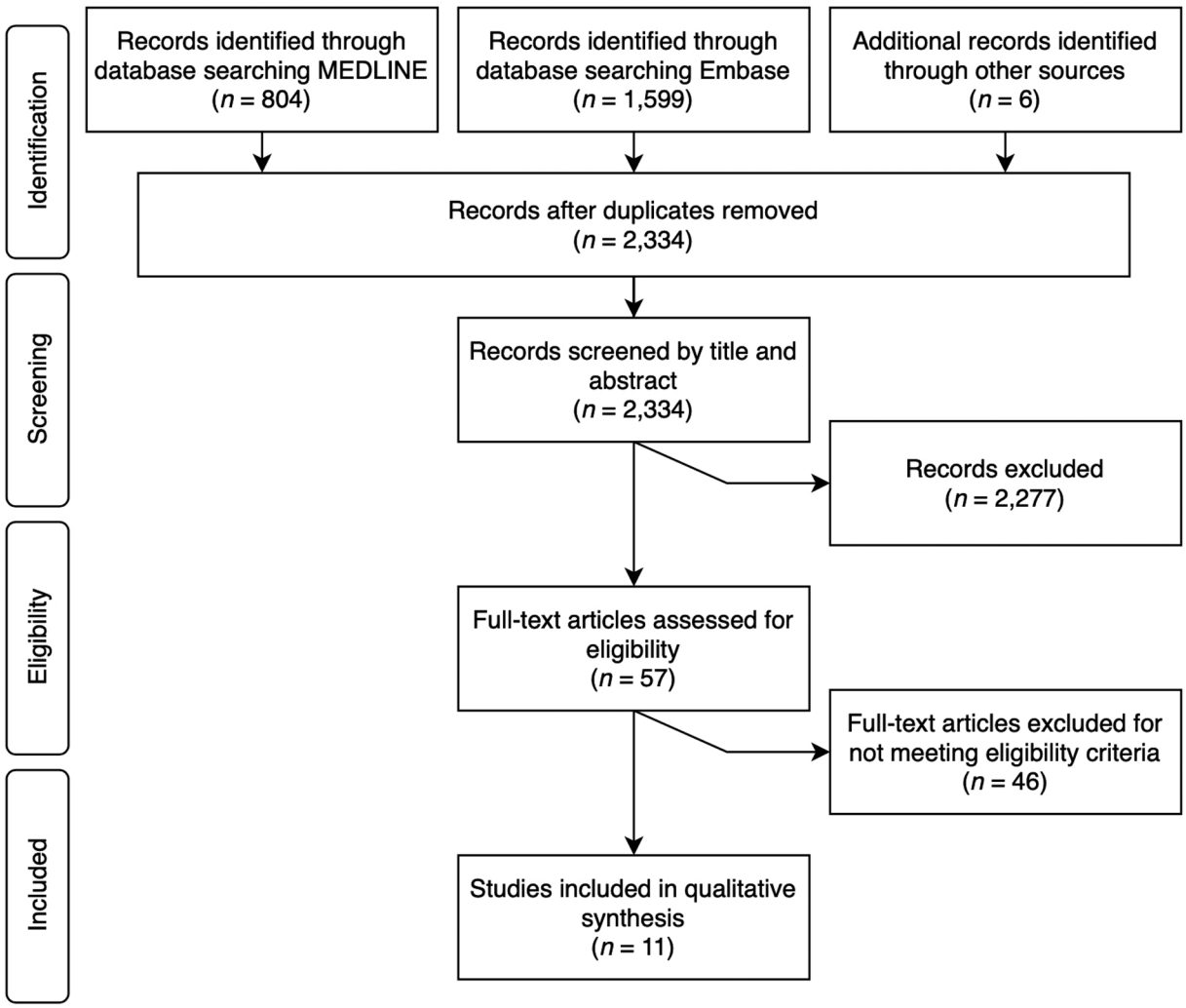

Fig. 1 PRISMA flow diagram of the study selection process.

Study Quality

Most models ($n = 7$) were developed using data from multiple centers, and the number of contributing centers ranged from 11 to 917. Two models were developed from single-center data. All models were developed using data on procedures performed between 1987 and 2014. Six of the nine models were developed from cohorts where no patient underwent intervention prior to 2000. Models were developed for all cardiac surgery procedures ($n = 2$), isolated CABG ($n = 5$), separate models for different procedure categories ($n = 1$), and for both CABG and percutaneous coronary intervention (PCI; $n = 1$).

Data missingness and handling was not reported in four of the nine of the studies. Strategies for missing data handling included complete case analysis ($n = 3$), multiple imputation ($n = 2$), and imputation of the median or mode ($n = 2$), with some studies employing more than one strategy depending on the variable. The ratio of events per variable (EPV) was not directly reported in two studies. EPV in studies where it could be calculated ranged from 10 to 141. All models were developed to predict mid- to long-term mortality ranging from 31 days to 7 years. One-year mortality was the most common outcome ($n = 4$).

Cox regression was used for model development in six studies, and logistic regression was used in three studies. Predictor selection techniques included full model approach ($n = 3$), backward stepwise selection ($n = 4$), stepwise selection where directionality was not reported ($n = 1$) or bootstrap bagging techniques ($n = 1$). Methods of internal validation included split sample ($n = 6$), cross-validation ($n = 2$), and bootstrapping ($n = 1$). Model discrimination was reported as an AUC or c-statistic in eight studies, ranging from 0.69 to 0.83. The AUCs of individual models within the studies and the relationship with follow-up time endpoints is shown in ►**Fig. 2**. A negative correlation between isolated CABG model AUCs and longer follow-up time endpoint was demonstrated and is shown in ►**Supplementary Fig. S1** (available in the online version). One study reported discrimination using the extreme quartile odds ratio. Calibration was reported using O:E ratios ($n = 4$), calibration curves ($n = 2$), the H–L test ($n = 1$), or not reported ($n = 2$).

MINORS score was 16 of 16 in two models, PROBAST risk of bias was low in three models, and applicability concern was

Table 1 Summary of studies

Author	Missing data method	Validation method	Development				Validation			
			n	EPV	Calibration	Discrimination	*n*	EPV	Calibration	Discrimination
Aktuerk et al[15]	Not reported	Split sample 70:30	Not directly reported	Not directly reported	Adequate	0.72–0.80	Not directly reported	Not directly reported	Adequate	0.69–0.79
Karim et al[16]	Multiple imputation	Split sample 50:50	11167–21258[a]	8-22[a]	Not reported	Not reported	123-304	9–23	Not reported	0.74–0.83
Tanaka et al [17]	Complete case analysis, multiple imputation	Cross-validation	Not directly reported	Not directly reported	Adequate	Not reported	8033	38	Adequate	0.79, 0.81
MacKenzie et al[18]	Complete case analysis	Cross-validation	13923-26919[a]	29-347[a]	Adequate	0.72–0.80	Not directly reported	Not directly reported	Not reported	Not reported
McDonald et al[24]	Complete case analysis	Bootstrapping	2031	24	Adequate	0.80	2031	24	Adequate	0.81
Shahian et al[20]	Single imputation	Split sample 50:50	174506	Not directly reported	Not reported	Not reported	173835	Not directly reported	Adequate	0.73–0.76
Toumpoulis et al[25]	Not reported	Split sample 67:33	3233	Not directly reported	Not reported	Not reported	1619	Not directly reported	Not reported	Extreme quartile odds ratio, 15.6
Wu et al[21]	Not reported	Split sample 50:50	Not directly reported	Not directly reported	Not reported	0.77–0.78	Not directly reported	Not directly reported	Adequate	0.77–0.78
Ziv-Baran et al[23]	Not reported	Split sample 50:50	1468	69	Adequate	0.74–0.76	1467	65	Adequate	0.76–0.77

Abbreviations: EPV, events per variable; *n*, number.
[a]Ranges stated for studies describing more than one model.

Thoracic and Cardiovascular Surgeon Vol. 72 No. 1/2024

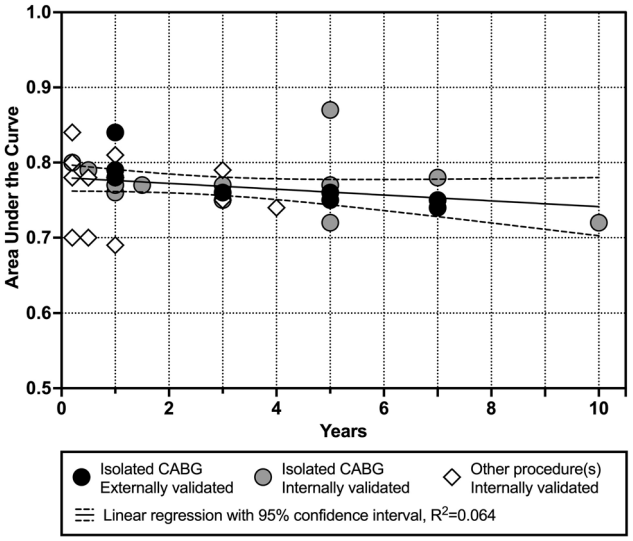

Fig. 2 Discrimination performance of internally and externally validated cardiac surgery clinical prediction models identified from the included studies for different long-term follow-up mortality endpoints. Linear regression line and 95% confidence intervals displayed.

low in eight models. These scores are detailed for each study in ►**Supplementary Table S3** (available online only).

Predictors

Final predictors included in the risk model were reported in all of the included studies. Models utilized preoperative predictors only ($n=7$) or preoperative and intra- or postoperative predictors ($n=2$). As shown in ►**Table 2**, the most frequently identified predictors were age ($n=9$), peripheral vascular disease (PVD; $n=8$), pulmonary disease ($n=8$), renal impairment ($n=8$), diabetes ($n=7$), ejection fraction ($n=7$), and congestive heart failure or New York Heart Association (NYHA) class ($n=6$). The number of predictors in the included risk models ranged from 9 to 32.

Preoperative Models

Aktuerk et al Model[15]

Data from 61,860 patients undergoing cardiac surgery in England between 2008 and 2011 at multiple centers were used to develop three separate models to predict outcomes for patients undergoing isolated CABG ($n=35{,}115$), isolated valve surgery ($n=1$ 8,353), and combined CABG and valve surgery ($n=8{,}392$). Endpoints were 90-day, 180-day, and 1-year mortality. Handling of missing data was not discussed. All three multivariable logistic regression models for isolated CABG were comprised of 15 predictors. The isolated valve models included 10 to 15 predictors, and all three combined CABG and valve models used 12 predictors. Internal validation was via a 70:30 split sample approach. AUCs for the combined cohort at 90 days, 180 days, and 365 days, respectively, were 0.80, 0.80, and 0.78 for CABG; 0.78, 0.78, and 0.78 for valve; and 0.74, 0.73, and 0.72 for combined CABG and valve. The models demonstrated adequate calibration (measured using calibration plots).

Karim et al Model[16]

This study included 46,573 patients undergoing isolated CABG in 28 centers across Australia between 2001 and 2014. Four separate Cox regression models were developed to predict mortality occurring within different time frames: 31–90 days, 91–365 days, 1–3 years, and >3 years. Missing data were handled using a multiple imputation approach. The four models were comprised of 13 predictors. Internal validation was performed with a 50:50 split sample approach and subsequent multifold cross-validation. AUCs for the four models in the validation set were 0.83, 0.79, 0.75, and 0.74, respectively. No measures of model calibration were discussed.

Kyoto et al Model[17]

The Kyoto model was developed to predict long-term outcomes for patients undergoing a first coronary intervention, either PCI or isolated CABG, between 2000 and 2002. The study cohort included 9393 patients, of whom 2,515 underwent surgery. Endpoints were 1 and 3-year mortality, and separate models were developed using logistic regression. A total of 19 predictors were included in the final model. Patients with missing data were excluded, with the exception of missing data points for ejection fraction, which was addressed using multiple imputation. AUCs for the model at 1 and 3-years were 0.81 and 0.79, respectively. Calibration was assessed using O:E ratios and was found to be acceptable at both endpoints. Analysis of model performance only for patients undergoing surgery was not reported.

MacKenzie et al Model[18]

The MacKenzie model was developed on data from 15,245 patients undergoing isolated CABG at several institutions across northern New England and United States between 1987 and 2001. Three separate Cox regression models were developed to predict mortality at different time periods: 0–3 months, 4–18 months, and >18 months. Patients with missing data were excluded, except for missing data for white cell count and body mass index (BMI), which were imputed using the mean value. The models comprised of 12 predictors. AUCs for the three models were 0.80, 0.77, and 0.72, respectively. Calibration, assessed using O:E ratios, was acceptable.

The MacKenzie model has also been externally validated in a study from Kilpin et al, which included 34,961 patients who underwent isolated CABG in 28 different centers across Australia since 2001.[19] Predicted mortality was compared with observed mortality at 1, 3, 5, and 7 years to assess model performance. AUCs were 0.79, 0.76, 0.76, and 0.75 for the four time points, respectively. Calibration was inadequate for all endpoints (all H–L test p-values <0.001), with the model significantly overpredicting mortality. The study concluded the MacKenzie model was not suitable for use in current Australian cardiac surgery practice.

Table 2 Individual model predictors

Variable	Aktuerk et al model[15]	Karim et al model[16]	Kyoto et al model[17]	MacKenzie et al model[18]	McDonald et al model[24]	Shahian et al model[20]	Toumpoulis et al model[25]	Wu et al model[21]	Ziv-Baran et al model[23]	Sum
Age	X	X	X	X	X	X	X	X	X	9
PVD	X	X	X	X		X	X	X	X	8
Pulmonary disease	X	X	X	X		X	X	X	X	8
Renal impairment	X	X	X	X		X	X	X	X	8
Diabetes		X	X	X		X	X	X	X	7
Ejection fraction		X	X	X		X	X	X	X	7
CHF or NYHA grade	X	X				X	X	X	X	6
Cerebrovascular disease		X	X		X	X		X		5
Arrhythmia		X	X			X		X		4
BMI			X	X		X		X		4
Gender			X	X	X	X				4
Operative urgency	X			X	X	X				4
Previous MI				X	X	X			X	4
IABP use					X	X			X	3
Left main stem disease			X			X		X		3
Smoking		X	X			X				3
Charlson score	X				X					2
Ethnicity						X	X			2
Haemodynamic instability or shock						X		X		2
Hyperlipidaemia		X	X							2
Hypertension		X				X				2
Liver disease	X		X							2
Malignancy	X		X							2
No. of coronary vessels involved				X		X				2

Table 2 (*Continued*)

Variable	Aktuerk et al model[15]	Karim et al model[16]	Kyoto et al model[17]	MacKenzie et al model[18]	McDonald et al model[24]	Shahian et al model[20]	Toumpoulis et al model[25]	Wu et al model[21]	Ziv-Baran et al model[23]	Sum
Previous cardiac surgery							X	X		2
Single model risk factor[a]	4	1	4	1	12	12	12	–	–	–
Total no. of predictors	10-15	13	19	12	21	32	21	13	9	

Abbreviations: BMI, body mass index; CABG, coronary artery bypass grafting; CHF, congestive heart failure; IABP, intra-aortic balloon pump; ICU, intensive care unit; MI, myocardial infarction; NYHA, New York Heart Association; PCI, percutaneous coronary intervention; PVD, peripheral vascular disease.

[a]The following risk factors were only used in one model: Anemia, aortic insufficiency, aortic stenosis, bilateral internal thoracic artery grafting, calcified aorta, carotid stenosis >75%, critical postoperative state, diseased coronary vessel, endocarditis, gastrointestinal complications, immune deficiency, immunosuppressive treatment, index of multiple deprivation, intraoperative stroke, mitral insufficiency, off-pump CABG, past smoker, PCI within 6-hours, preoperative ICU stay, PCI in past year, previous sternotomy, postoperative readmission to ICU, postoperative cardiac reopening, postoperative dialysis, postoperative massive transfusion, postoperative multiple organ dysfunction, postoperative respiratory failure, postoperative vasoactive and inotropic medications, postoperative ventilation, procedure, proximal left anterior descending artery disease, proximal left anterior descending artery total occlusion, preoperative IABP/inotropes, previous PCI, prior carotid surgery, recent admission, recent emergency admission, reoperation, sepsis and/or endocarditis, thrombolysis prior to operation, tricuspid insufficiency, steroid use, urgency of admission, white blood cell count, and year of surgery.

Shahian et al Model[20]

This model was developed using data from 348,341 patients aged 65 years and above undergoing isolated CABG across multiple American centers between 2002 and 2007. Three separate Cox regression models were constructed to predict postoperative mortality between 31 and 180 days, 181 days and 2 years, and >2 years after surgery. A 50:50 split sample approach was used for internal validation. Missing data were replaced using the median and most common values for continuous and noncontinuous data, respectively. The models contained 32 different predictors. Survival was dichotomized at 1 and 3 years for which AUCs were 0.76 and 0.75, respectively. Calibration, measured using O:E ratios, was acceptable.

Wu et al Model[21]

This model was developed from 8,597 patients undergoing isolated CABG during July to December 2000 in 33 hospitals across the New York state. The model was constructed using Cox regression. Outcomes were mortality at 1, 3, 5, and 7 years after surgery. Handling of missing data was not reported. Internal validation was performed using a 50:50 split sample approach. A total of 13 predictors were retained for inclusion in the final model. The AUCs for 1, 3, 5, and 7-year mortality were 0.77, 0.77, 0.77, and 0.78, respectively. Calibration, assessed using O:E ratios, was acceptable.

The Wu model has been externally validated in two separate studies. Carr et al externally validated the model in 1,028 patients undergoing isolated CABG in a single North American center between 2006 and 2011.[22] AUCs for 1, 3, and 5-year mortality were 0.84, 0.76, and 0.76, respectively. Calibration, which was assessed using both O:E ratios and the H–L test, was also acceptable.

Kilpin et al also externally validated the model in the same study that validated the MacKenzie model.[19] Measured endpoints were 1, 3, 5, and 7-year mortality. Discrimination was 0.78, 0.76, 0.75, and 0.74 for the four endpoints. Calibration was inadequate for all endpoints (all p-values <0.001), with the model significantly overpredicting mortality. The study concluded that the Wu model was not suitable for use in current Australian cardiac surgery practice.

Ziv-Baran et al Model[23]

A cohort of 2,935 patients undergoing isolated CABG with bilateral internal mammary artery grafting in a single center in Israel between 1996 and 2011 was used to create a nomogram to predict 5, 10, and 15-year mortality. The model was developed using Cox regression and included nine predictors. Handling of missing data was not discussed. Internal validation was performed using a 50:50 split-sample approach. AUCs for the three endpoints were 0.87, 0.72, and 0.53, respectively. Calibration was acceptable for all three endpoints and was assessed using calibration curves.

Models Including Intra- or Postoperative Predictors

McDonald et al Model[24]

This model was derived using data from 2,031 patients undergoing first-time cardiac surgery in 11 different centers

across Ontario, Canada, between 2009 and 2014. Only patients who experienced a prolonged critical care stay (defined as 7 days or more) were included. The primary outcome was death between 7 and 365 days after surgery. The model was developed using multivariable logistic regression and was internally validated using a bootstrapping approach. A total of 21 predictors were included in the final model. Intra- and postoperative predictors included in the model were readmission to intensive care unit (ICU), reoperation, intra-aortic balloon pump use, hemofiltration, vasoactive or inotrope medication use, ventilation, massive transfusion, and multiple organ dysfunction. Patients with missing data were excluded. The model AUC was 0.80 (0.81 after bootstrapping), and calibration was acceptable according to the H–L test ($p = 0.500$).

Toumpoulis et al Model[25]

The Toumpoulis model was developed using data from 4,852 patients undergoing cardiac surgery in a single North American center between 1992 and 2002. It was designed to predict 5-year survival and was developed using multivariable Cox regression. Subsequent internal validation was undertaken using a 67:33 split sample approach. Intra- or postoperative predictors included were intraoperative stroke, critical postoperative state, sepsis and/or endocarditis, gastrointestinal complications, and respiratory failure. Handling of missing data was not discussed. Model performance was assessed by dividing patients into four quartiles according to predicted risk and analyzing estimated survival using the Kaplan–Meier method. Overall survival was significantly different between the four quartiles ($p < 0.001$). No other measures of discrimination or calibration were reported.

Discussion

This systematic review identified nine clinical prediction models for long-term outcomes after cardiac surgery. Two studies undertaking external validations of two of these models were also identified. Predictors included in the models were relatively homogenous between models despite a large number of predictors included across the studies and a range of timeframes for the outcomes. Model performance measures largely demonstrated acceptable discrimination and calibration on internal validation, with performance appearing to diminish over time. However, 95% confidence intervals for AUC or calibration were not reported in three of the studies. The use of AUC as the sole measure of model performance (as was the case in two of the studies included in this systematic review) is insufficient.[26]

Seven models included solely preoperative predictors, and two included intra- and postoperative data. The timing of predictor measurement is an important consideration for clinical prediction model development. A model designed to aid clinical decision-making regarding the proposed procedure can clearly only include preoperative data. Most models were either designed with the potential to inform preoperative clinical decision-making or did not specify a clear application. Models designed to predict long-term outcomes are unlikely to be heavily relied upon for preoperative counseling unless a procedure is indicated solely for the prognostic benefit or for comparing competing treatment options.[17,27]

An important function of long-term clinical prediction modeling is likely to be facilitating comparative clinical governance analyses by allowing risk adjustment. A further application may be informing patients about individualized prognosis during the postoperative period. For these latter two purposes, intra- and postoperative data may be useful in addition to baseline risk. Some intra- and postoperative predictors identified in this review include cardiopulmonary bypass use, ICU readmission, postoperative complications, and reoperation. Given their relative infrequency as predictors in long-term risk models, identifying important intra- or postoperative parameters based on this systematic review remains challenging.

Before a model is applied in clinical practice, it should ideally be externally validated in a cohort similar to that in which it is intended to be used.[28,29] Only two of the identified models have undergone external validation to date. External validation is an important step as internal validation may overestimate model performance and may not be sufficient in assessing a model's generalisability.[30,31] External validation can also help identify the need for model updating or recalibration. The external validation study testing the MacKenzie et al model in an Australian population suggested the model was poorly calibrated and overpredicted mortality risk.[19] The development of the MacKenzie et al model included patients undergoing surgery from 20 to 34 years ago. Significant calibration drift has previously been demonstrated in cardiac surgery for in-hospital mortality, and the overprediction of risk by the MacKenzie et al model again highlights the importance of considering the time that has elapsed between model development and clinical use.[32]

The acceptable performance of the majority of models on internal validation suggests using existing databases may be a reasonable approach in developing a long-term risk model. However, the potential benefits of including other variables more suitable in the prediction of longer-term events, for instance by building a new database, must also be considered. Ultimately, a balance needs to be struck whereby predictors must be available, reliably measured, and occur with enough frequency within the population to be both relevant and valid given the intended use of the model.

Both prospective clinical studies and retrospective observational studies can be utilized for clinical prediction modeling. Prospective studies may be considered preferential due to their typically superior measurement and reporting standards of predictors and outcomes.[33] However, studies using retrospective datasets or clinical registries are more likely to be powered to detect differences in clinical endpoints as required for clinical governance purposes, rather than for model development alone.[34,35] Larger studies retrospectively analyzing databases of prospectively collected data, such as the Karim, MacKenzie, or Shahian et al models link to national death indices to measure their endpoint: all-

cause mortality.[16,18,21] This facilitates a significantly larger sample size and, consequently, the potential for the inclusion of a greater number of candidate predictors.

Although long-term all-cause mortality is a useful outcome that can be addressed using existing registries or databases, existing databases may not include other outcome measures of treatment efficacy such as symptomatic relief, functional status, or freedom from reintervention. No models predicting long-term quality of life measures were identified, despite these being important long-term outcomes alongside measures of safety such as mortality.[36–38] Although patient-reported outcomes would be very difficult to collect as part of existing clinical registries, there is scope through more extensive registry linkage to identify composite end-points that reflect the efficacy of treatment.

No studies identified in this review were developed for a composite outcome measure. Composite end-points have been used in contemporary high-profile clinical trials such as EXCEL, NOBLE, PRECOMBAT, and SYNTAX.[39–42] Utilizing carefully considered composite outcomes, such as major adverse cardiac or cerebrovascular events, would increase the number of events for modeling studies. Guidance exists on which standardized endpoints should be measured in clinical effectiveness trials, such as that published by the Academic Research Consortium and the Valve Academic Research Consortium.[43,44] However, these recommendations are procedure specific and designed for prospectively conducted clinical studies rather than prediction model development studies. Identifying composite outcomes that are clinically relevant, available, and appropriate for both prospective clinical studies and model development studies is an area where further work is required.

As demonstrated in this review, long-term risk prediction modeling in cardiac surgery thus far has focused primarily on mortality. Though all models identified in this review focus on mortality, there is no consistency with regard to the time frame. Most models have been developed for 1-year mortality, and only one assessed mortality more than 5 years after surgery. The early risk of mortality after cardiac surgery has been suggested to continue for up to approximately 120 days, suggesting that outcomes such as 1-year mortality do not adequately represent the long-term efficacy of treatment.[45] For an outcome to be considered for wider use, standardization of an appropriate long-term outcome timeframe post-cardiac surgery would be valuable.

Limitations of this systematic review include the potential incomplete retrieval of relevant research due to limiting the inclusion of studies written in English. This review also only included studies presenting full risk models and excluded studies assessing associations between risk factors and long-term outcomes. As such, information on potentially useful predictors of long-term outcomes may have been omitted. Additionally, models developed to predict short-term outcomes that have been validated for long-term outcomes were excluded as they were beyond the scope of this study, though this topic does warrant further study given their potential for the prediction of long-term outcomes.[23,46–54]

Conclusion

This review has identified nine risk prediction models developed to predict long-term mortality after cardiac surgery. The included predictors used were largely preoperative, which is useful for conventional risk adjustment and informing patients of risk preoperatively. Some models that include intra- and postoperative variables were identified, and this approach could potentially be useful for comparative clinical governance analyses. All models performed well on internal validation, but only two models were validated externally. Both performed poorly on external validation, and only one study validated the models for greater than 5 years of follow-up. No model identified in this systematic review can, therefore, be recommended for use in contemporary cardiac surgical practice. Before long-term risk modeling after cardiac surgery can be implemented, the development of new models using a standardized composite outcome measure or successful external validation of existing models is required.

Authors' Contribution

L.A., M.T., and S.W.G. designed the study. L.A. and M.T. collected the data. L.A., M.T., and S.W.G analyzed and interpreted the data. L.A.,M.T., M.O., R.V. and S.W.G. wrote the manuscript.

Conflict of Interest

None declared.

References

1 Bridgewater BSociety for Cardiothoracic Surgery in GB and Ireland. Cardiac registers: the adult cardiac surgery register. Heart 2010;96(18):1441–1443

2 Grant SW, Kendall S, Goodwin AT, et al. Trends and outcomes for cardiac surgery in the United Kingdom from 2002 to 2016. JTCVS Open 2021;7:259–269

3 National Institute for Cardiovascular Outcomes Research (NICOR) National Adult Cardiac Surgery Audit (NACSA) 2020 Summary Report 2016/17–2018/19 Data. (2020). Accessed June 01, 2021 at: https://www.nicor.org.uk/wp-content/uploads/2020/12/National-Adult-Cardiac-Surgery-Audit-NACSA-FINAL.pdf

4 Leon MB, Smith CR, Mack M, et al; PARTNER Trial Investigators. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. N Engl J Med 2010;363(17):1597–1607

5 Adams DH, Popma JJ, Reardon MJ, et al; U.S. CoreValve Clinical Investigators. Transcatheter aortic-valve replacement with a self-expanding prosthesis. N Engl J Med 2014;370(19):1790–1798

6 Stone GW, Lindenfeld J, Abraham WT, et al; COAPT Investigators. Transcatheter mitral-valve repair in patients with heart failure. N Engl J Med 2018;379(24):2307–2318

7 Nashef SAM, Roques F, Michel P, Gauducheau E, Lemeshow S, Salamon R. European System for Cardiac Operative Risk Evaluation (EuroSCORE). Eur J Cardiothorac Surg 1999;16(01):9–13

8 Nashef SAM, Roques F, Sharples LD, et al. EuroSCORE II. Eur J Cardiothorac Surg 2012;41(04):734–744, discussion 744–745

9 Jin R, Furnary AP, Fine SC, Blackstone EH, Grunkemeier GL. Using Society of Thoracic Surgeons risk models for risk-adjusting cardiac surgery results. Ann Thorac Surg 2010;89(03):677–682

10 Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PLoS One 2012;7(02):e32844

11 Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014;11 (10):e1001744
12 Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162(01):W1-73
13 Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (minors): development and validation of a new instrument. ANZ J Surg 2003;73(09):712–716
14 Wolff RF, Moons KGM, Riley RD, et al; PROBAST Group† PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019;170(01):51–58
15 Aktuerk D, McNulty D, Ray D, et al. National administrative data produces an accurate and stable risk prediction model for short-term and 1-year mortality following cardiac surgery. Int J Cardiol 2016;203:196–203
16 Karim MN, Reid CM, Huq M, et al. Predicting long-term survival after coronary artery bypass graft surgery. Interact Cardiovasc Thorac Surg 2018;26(02):257–263
17 Tanaka S, Sakata R, Marui A, Furukawa Y, Kita T, Kimura TCREDO-Kyoto Investigators. Predicting long-term mortality after first coronary revascularization: – the Kyoto model –. Circ J 2012;76(02): 328–334
18 MacKenzie TA, Malenka DJ, Olmstead EM, et al; Northern New England Cardiovascular Disease Study Group. Prediction of survival after coronary revascularization: modeling short-term, mid-term, and long-term survival. Ann Thorac Surg 2009;87(02):463–472
19 Kilpin M, Talwar A, Meneguzzi J, Tran L, Reid C, Hayward P. Two long-term mortality risk models for coronary artery bypass graft surgery produced in American Populations validated in an Australian population. Heart Lung Circ 2018;27(01):79–88
20 Shahian DM, O'Brien SM, Sheng S, et al. Predictors of long-term survival after coronary artery bypass grafting surgery: results from the Society of Thoracic Surgeons Adult Cardiac Surgery Database (the ASCERT study). Circulation 2012;125(12):1491–1500
21 Wu C, Camacho FT, Wechsler AS, et al. Risk score for predicting long-term mortality after coronary artery bypass graft surgery. Circulation 2012;125(20):2423–2430
22 Carr BM, Romeiser J, Ruan J, et al. Long-term post-CABG survival: performance of clinical risk models versus actuarial predictions. J Card Surg 2016;31(01):23–30
23 Ziv-Baran T, Mohr R, Pevni D, Ben-Gal Y. A simple-to-use nomogram to predict long term survival of patients undergoing coronary artery bypass grafting (CABG) using bilateral internal thoracic artery grafting technique. PLoS One 2019;14(10):e0224310
24 McDonald B, van Walraven C, McIsaac DI. Predicting 1-year mortality after cardiac surgery complicated by prolonged critical illness: derivation and validation of a population-based risk model. J Cardiothorac Vasc Anesth 2020;34(10):2628–2637
25 Toumpoulis IK, Anagnostopoulos CE, Ioannidis JP, et al. The importance of independent risk-factors for long-term mortality prediction after cardiac surgery. Eur J Clin Invest 2006;36(09): 599–607
26 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010;21(01):128–138
27 Grant SW, Grayson AD, Mitchell DC, McCollum CN. Evaluation of five risk prediction models for elective abdominal aortic aneurysm repair using the UK National Vascular Database. Br J Surg 2012;99(05):673–679
28 Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 2012;98(09):691–698
29 Debray TPA, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 2015;68(03):279–289
30 Bleeker SE, Moll HA, Steyerberg EW, et al. External validation is necessary in prediction research: a clinical example. J Clin Epidemiol 2003;56(09):826–832
31 Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016;69:245–247
32 Hickey GL, Grant SW, Murphy GJ, et al. Dynamic trends in cardiac surgery: why the logistic EuroSCORE is no longer suitable for contemporary cardiac surgery and implications for future risk models. Eur J Cardiothorac Surg 2013;43(06):1146–1152
33 Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009;338:b375
34 Chen L. Overview of clinical prediction models. Ann Transl Med 2020;8(04):71
35 Hickey GL, Grant SW, Cosgriff R, et al. Clinical registries: governance, management, analysis and applications. Eur J Cardiothorac Surg 2013;44(04):605–614
36 Peric V, Stolic R, Jovanovic A, et al. Predictors of quality of life improvement after 2 years of coronary artery bypass surgery. Ann Thorac Cardiovasc Surg 2017;23(05):233–238
37 Herlitz J, Brandrup-Wognsen G, Caidahl K, et al. Determinants for an impaired quality of life 10 years after coronary artery bypass surgery. Int J Cardiol 2005;98(03):447–452
38 Pačarić S, Turk T, Erić I, et al. Assessment of the quality of life in patients before and after coronary artery bypass grafting (CABG): a prospective study. Int J Environ Res Public Health 2020;17(04):1417
39 Stone GW, Kappetein AP, Sabik JF, et al; EXCEL Trial Investigators. Five-year outcomes after PCI or CABG for left main coronary disease. N Engl J Med 2019;381(19):1820–1830
40 Head SJ, Davierwala PM, Serruys PW, et al. Coronary artery bypass grafting vs. percutaneous coronary intervention for patients with three-vessel disease: final five-year follow-up of the SYNTAX trial. Eur Heart J 2014;35(40):2821–2830
41 Holm NR, Mäkikallio T, Lindsay MM, et al; NOBLE investigators. Percutaneous coronary angioplasty versus coronary artery bypass grafting in the treatment of unprotected left main stenosis: updated 5-year outcomes from the randomised, non-inferiority NOBLE trial. Lancet 2020;395(10219):191–199
42 Ahn J-M, Roh J-H, Kim Y-H, et al. Randomized trial of stents versus bypass surgery for left main coronary artery disease: 5-year outcomes of the PRECOMBAT study. J Am Coll Cardiol 2015;65(20): 2198–2206
43 Garcia-Garcia HM, McFadden EP, Farb A, et al; Academic Research Consortium. Standardized end point definitions for coronary intervention trials: the Academic Research Consortium-2 Consensus Document. Circulation 2018;137(24):2635–2650
44 Kappetein AP, Head SJ, Généreux P, et al; Valve Academic Research Consortium-2. Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic Research Consortium-2 consensus document. J Thorac Cardiovasc Surg 2013;145(01):6–23
45 Siregar S, Groenwold RHH, de Mol BA, et al. Evaluation of cardiac surgery mortality rates: 30-day mortality or longer follow-up? Eur J Cardiothorac Surg 2013;44(05):875–883
46 Puskas JD, Kilgo PD, Thourani VH, et al. The society of thoracic surgeons 30-day predicted risk of mortality score also predicts long-term survival. Ann Thorac Surg 2012;93(01):26–33, discussion 33–35
47 Nilsson J, Algotsson L, Höglund P, Lührs C, Brandt J. Comparison of 19 pre-operative risk stratification models in open-heart surgery. Eur Heart J 2006;27(07):867–874
48 Habib AM, Dhanji AR, Mansour SA, Wood A, Awad WI. The EuroSCORE: a neglected measure of medium-term survival following cardiac surgery. Interact Cardiovasc Thorac Surg 2015;21 (04):427–434

49 Nonaka M, Komiya T, Shimamoto T, Matsuo T. Comparison of clinical outcomes after coronary artery bypass grafting using stratified SYNTAX scores. Gen Thorac Cardiovasc Surg 2020;68 (11):1270–1277
50 Sotomi Y, Cavalcante R, van Klaveren D, et al. Individual long-term mortality prediction following either coronary stenting or bypass surgery in patients with multivessel and/or unprotected left main disease: an external validation of the SYNTAX score II model in the 1,480 patients of the BEST and PRECOMBAT randomized controlled trials. JACC Cardiovasc Interv 2016;9(15):1564–1572
51 Hoogerduijn JG, de Rooij SE, Grobbee DE, Schuurmans MJ. Predicting functional decline in older patients undergoing cardiac surgery. Age Ageing 2014;43(02):218–221
52 Paul M, Raz A, Leibovici L, Madar H, Holinger R, Rubinovitch B. Sternal wound infection after coronary artery bypass graft surgery: validation of existing risk scores. J Thorac Cardiovasc Surg 2007;133(02):397–403
53 Biancari F, Asim Mahar MA, Kangasniemi OP. CHADS2 and CHA2DS2-VASc scores for prediction of immediate and late stroke after coronary artery bypass graft surgery. J Stroke Cerebrovasc Dis 2013;22(08):1304–1311
54 Sündermann S, Dademasch A, Rastan A, et al. One-year follow-up of patients undergoing elective cardiac surgery assessed with the comprehensive assessment of frailty test and its simplified form. Interact Cardiovasc Thorac Surg 2011;13(02):119–123, discussion 123