Risk models that utilise postoperative patient monitoring data to predict outcomes in adult cardiac surgery; a systematic review.

DOI:
10.1053/j.jvca.2016.10.002

Citation for published version (APA):
Howitt, S., Grant, S., Riding, D., Malagon, I., & Mccollum, C. (2017). Risk models that utilise postoperative patient monitoring data to predict outcomes in adult cardiac surgery; a systematic review. *Journal of Cardiothoracic and Vascular Anesthesia*, *31*(5). https://doi.org/10.1053/j.jvca.2016.10.002

Published in:
Journal of Cardiothoracic and Vascular Anesthesia

Risk models that utilise postoperative patient monitoring data to predict outcomes in adult cardiac surgery; a systematic review

Howitt SH, [1,2] Grant SW[1,3] Riding DM, [1] Malagon I,[2] McCollum CN, [1]

1) Academic Surgery Unit, Institute of Cardiovascular Sciences, University of Manchester, University Hospital of South Manchester, Manchester, UK.

2) Department of Cardiothoracic Anaesthesia and Critical Care, University Hospital of South Manchester, Manchester, UK.

3) National Institute for Cardiovascular Outcomes Research, University College London, Institute of Cardiovascular Science, London, UK

ADDRESS FOR CORRESPONDENCE

Prof Charles McCollum

Academic Surgery Unit, Education & Research Centre

University Hospital of South Manchester

Southmoor Road, Manchester, UK, M23 9LT

Telephone - +44 (0)161 291 5842

Fax - +44 (0)161 291 5854,

E-mail: cnmcc@manchester.ac.uk

ARTICLE TYPE: Review

KEYWORDS: Intensive Care; Cardiac Surgery; Complications; risk prediction models

WORD COUNT: 5417

*Revised Manuscript

ABSTRACT

Background

Preoperative risk prediction models are used to provide patients with information on perioperative mortality and to risk-adjust surgical outcome analyses. However, risk estimates from preoperative models may become increasingly unrealistic after surgery as they cannot take into account postoperative events. A number of risk models that utilise postoperative data have been developed or validated for adult cardiac surgery but none has been widely adopted. The objective of this review was to identify all such risk prediction models and discuss their uses and limitations.

Methods

A systematic review of the literature was undertaken with Medline, EMBASE and the Cochrane Library searched to identify relevant papers. Identified studies were assessed with regards to model discrimination, model calibration and clinical validity.

Results

The search identified 1649 publications. 86 met the inclusion criteria from which 14 validated models were identified. Eight models were originally designed for use in general intensive care units but subsequently validated for use following cardiac surgery. Six models were designed specifically for cardiac surgery patients. Most models that demonstrated good statistical performance were designed for clinical benchmarking purposes. No validated model provides predictions for specific complications or patient deterioration more frequently than once daily.

Conclusions

This review has identified a number of risk prediction models that utilise postoperative data and have been validated for the prediction of outcomes after adult cardiac surgery. The lack of adoption of these models may be due to variations in patient monitoring protocols and the inability of existing models to guide clinical decision making for individual patients. The risk scores identified are likely to be useful for assessing Cardiac Intensive Care Unit performance, informing discussions with patients or relatives and allocating resources. Future research to develop and validate predictive models that utilise postoperative data to produce frequent estimates of risk for specific patient outcomes may be of benefit.

INTRODUCTION

The most commonly used risk prediction tools in European adult cardiac surgery are the EuroSCORE models.[1,2] These models use preoperative patient data to predict postoperative mortality. They play a vital role in preoperative clinical decision making, informed consent and performance monitoring. However, they have limited clinical value in subsequent patient management as the predicted risk cannot be modified by the occurrence of significant postoperative events or the patient's response to those events. Consequently, risk estimates may become unrealistic as postoperative events unfold.

Currently, adult cardiac surgery carries a mortality risk of 2-3% [3,4]. This risk is significantly higher in those who develop postoperative complications. Respiratory [5,6] and renal failure [7,8] following cardiac surgery are associated with mortality rates of up to 18% and 60% respectively. Models that identify patients at risk of such complications could reduce morbidity and mortality by alerting clinicians to those who would benefit from early, targeted interventions.

A number of risk prediction models that utilise postoperative data have been developed or validated for use in adult cardiac surgery. Some models calculate risk based on the initiation of treatments or the occurrence of events in the postoperative period.[9-12] These models may provide updated risk estimates that guide staff and resource allocation and may also inform discussions with patients and their relatives. However, they often only demonstrate increased risk once end organ damage has occurred and remedial measures have been taken. Accordingly, they are of limited use in the early identification of those at risk and may not enable timely administration of preventative treatment. Their usefulness for benchmarking may be limited by interinstitutional variation in initiation of treatments according to local protocols. Models based on postoperative physiological monitoring data are potentially better suited to these tasks. Such models share similarities with Early Warning Scores (EWS)[13], which have been widely adopted to identify ward-based patients at risk of clinical deterioration based on analyses of physiological values including heart rate, respiratory rate, oxygen saturation, blood pressure, temperature and conscious level. Despite widespread adoption of EWS models on other wards and the availability of vast amounts of patient monitoring data in the ICU setting following cardiac surgery, no risk model based on patient monitoring data following cardiac surgery has been widely adopted. The objective of this review was to identify all validated risk models which use postoperative patient monitoring data to predict outcomes in adult cardiac surgery. Clinical validity and statistical performance were evaluated to explore possible reasons for the lack of adoption.

METHODS

Literature search and study eligibility

The Database of Abstract of Reviews of Effects (DARE) and PubMed Health databases were searched using the terms “cardiac surgery” or “coronary artery bypass” or “valve “and “risk prediction” or “model” for papers published since 2009 and revealed no existing Cochrane, CRD or PubMed Health registered reviews. A subsequent search of the Cochrane library, EMBASE and MEDLINE databases from inception to 2015 was performed using the PICOS framework (Appendix A). Two "readers" (SHH and DMR) independently screened the titles and abstracts to select potentially eligible studies. The full text of potentially eligible manuscripts was assessed by both readers independently. Studies were eligible if they reported the validation of a risk prediction model using postoperative patient monitoring data to predict outcomes after adult cardiac surgery. In addition to the validation study, the article that first described the validated model was identified and reviewed for details concerning model development. There were no restrictions on study design. Only studies presented in English were analysed.

Data extraction and quality assessment

Data was extracted from the eligible manuscripts by SHH and included first author’s name, year of publication, study design, sample size and population characteristics. For studies describing the development of a risk prediction model information extracted included; statistical model used, factors included in the model, model outcomes and method of validation. For articles describing the validation (internal or external) of a risk prediction model in cardiac surgery patients information extracted included; the quality of the study, statistical performance of the model and characteristics of the validation cohort.

When assessing the models, three main aspects of their performance were considered: discrimination, calibration and clinical validity. Discrimination was usually assessed using the area under the Receiver Operator Characteristic curve (AUC).[14] An AUC of 0.5 represents discrimination between patients who experience an outcome and those who do not, that is no better than chance. An AUC of 1.0 represents perfect discrimination, with values >0.7 generally accepted to indicate adequate discrimination, and >0.8 considered good.[15-17]

Calibration, or how closely the predicted risk matches the observed risk, can be assessed using a variety of different methods. The Hosmer-Lemeshow (HL) test was most commonly used. A high HL χ^2 value with a low associated p value suggests that there is a significant difference between predicted risk and observed outcomes across sub-groups of the cohort.[18] Other calibration measures included the Brier and the R^2 score. Brier score values approaching zero represent good calibration. The R^2 score is used for continuous outcomes e.g. length of stay, with a value of 1 indicating perfect fit. Clinical validity was assessed considering the quality of the study design, the methodology and the reporting.

RESULTS

A total of 86 relevant studies were identified (Figure 1). Amongst these there were 14 risk models which had been validated for use in cardiac surgery patients (Table 1). Eight of these models were initially developed using data from general ICU populations with half of these developed using cohorts from which cardiac surgery patients were excluded. Six models were developed using only patients who had undergone cardiac surgery. Most models were developed using logistic regression but expert opinion, Bayesian modelling and Gaussian processes were also utilised. (Table 1)

The overall quality of these studies was good (Table 5). The main limitation was a failure to clearly describe how missing data was handled. Occasionally, preoperative patient characteristics were not included, but in these studies composite measures of patient co-morbidity such as the mean Euroscore were usually provided.

Five of the 14 models included purely postoperative variables, four included preoperative and postoperative variables and five models included intraoperative, preoperative and postoperative variables. The variables used by the validated models are detailed in Table 2. The organ system most commonly assessed using patient monitoring data was the cardiovascular system. Many models simply include the mean arterial pressure while some depend on knowledge of cardiac output. Others use the composite measure of Pressure Adjusted Heart Rate which is based on heart rate, central venous pressure and mean arterial pressure. The respiratory system was most commonly assessed using the ratio of arterial partial pressure of oxygen to inspired oxygen concentration. The renal system was assessed using blood test results rather than urine output in all but four models. Temperature was measured in five models.

The statistical performance of the ten models validated for prediction of mortality is shown in Table 3. The statistical performance of the models validated for the prediction of morbidity is shown in Table 4. Morbidity outcomes predicted included prolonged ICU stay, prolonged ventilation, acute kidney injury (AKI) and composite morbidity. A number of models were developed and validated for both mortality and morbidity. APACHE-II, SAPS-II, SOFA, ICURS and CASUS were validated in multiple patient cohorts. These all showed good discrimination in multiple studies with ROCs > 0.75. Of those validated in multiple studies, SOFA and CASUS scores consistently demonstrated the best combinations of AUCs >0.8 and p values > 0.05 for the HL χ^2 test in external validation cohorts.

Models developed for general ICU and validated in cardiac surgery patients

Acute Physiology and Chronic Health Evaluation II (APACHE II)

In 1985 Knaus et al. developed the APACHE II score[19] from the original APACHE score.[20] APACHE-II estimates the risk of mortality for ICU patients using data including patient age, co-morbidity and an Acute Physiologic Score (APS) based on the most abnormal values of 12 physiological variables recorded during the first 24 hours of ICU admission. Cardiac surgery patients were excluded from the model's development dataset.

A 2001 study by Kern et al. demonstrated that APACHE-II discriminated well when predicting prolonged mechanical ventilation in 687 cardiac surgery patients.[21] In 2005 Hekmat et al. demonstrated that APACHE-II scores calculated daily for 1057 cardiac surgery patients performed well, with postoperative day 3 scores best predicting 30 day mortality.[22] In 2011, Doerr et al. conducted similar analyses using the records of 2801 cardiac surgery patients.[23] When predicting ICU mortality APACHE-II showed adequate discrimination for each postoperative day but calibration was only adequate on the first postoperative day. Mean and worst APACHE-II scores for each patient were also used to generate mortality predictions with the mean APACHE-II score showing best discrimination and calibration.

Exarchopoulos and colleagues demonstrated that APACHE-II scores at ICU admission successfully predicted 30 day mortality in 150 cardiac surgery patients.[24] Similarly Tsaousi et al. demonstrated that ICU admission APACHE-II score successfully predicted in-hospital mortality in 1058 cardiac surgery patients.[25] However, in a UK study, Ariyaratnam et al. found admission APACHE-II scores poorly predicted perioperative mortality.[3]

Acute Physiology and Chronic Health Evaluation III (APACHE-III)

APACHE-III was developed using data from 17,440 patients from 40 hospitals.[26] The same physiological variables included in APACHE-II were measured in the first 24 hours of admission, together with urine output and four additional blood analyses. The final model included 17 physiological variables which combined to create the APS. Compared with APACHE-II, APACHE-III assigns greater weight to extremely abnormal values. The APS is combined with chronic disease status and age to produce the final APACHE-III score. As with APACHE-II, cardiac surgery patients were not included in the development cohort.

A model including APS from APACHE-III, patient information and surgery type was validated in 2435 coronary artery bypass graft (CABG) patients.[27] This discriminated well when predicting hospital mortality for groups of patients, but in individuals the APS scores correlated poorly with mortality, length of ICU stay and treatment costs.

Simplified Acute Physiology Score II (SAPS-II)

The SAPS-II model was developed using data from 137 centres across 12 countries over a six month period in 1991-1992.[28] SAPS-II was designed for general ICUs. Cardiac surgery patients were excluded. Similarly to the APACHE scores, this model also used the worst recorded value for each variable during the first 24 hours of admission.

The ability of daily SAPS-II scores to predict 30-day mortality after cardiac surgery was also assessed in Doerr's 2011 study.[23] Discrimination was found to be good but the model was poorly calibrated in this group of patients. Derived variables such as maximum and mean SAPS–II score showed excellent discrimination and calibration. The same author subsequently analysed mortality predictions for 5207 cardiac surgery patients (including the initial 2801). The calibration of daily SAPS-II scores was inadequate, but again discrimination was acceptable.[29]

Kern et al. also assessed the ability SAPS-II to predict prolonged mechanical ventilation after cardiac surgery, reporting good discrimination but without commenting on calibration.[12] Exarchopoulos et al. found that admission SAPS-II score performed well when predicting 30 day mortality in a study of 150 cardiac surgery patients.[24]

Multiple Organ Dysfunction Score (MODS)

In 1995 Marshall et al. described the MODS as a tool to grade the severity of organ dysfunction in patients admitted to a Canadian surgical ICU between 1988 and 1990.[30] The score was developed in order to measure patients' progress on a daily basis during ICU stay and used data from 336 patients to grade dysfunction in 6 major organ systems.

In 2005 and 2010 Hekmat et al. published validation studies in which MODS was calculated daily in two cohorts of 384 and 1057 cardiac surgery patients.[22,31] MODS had good discriminatory abilities with some variation depending on the day on which the score was calculated. Calibration was reported as acceptable, although p values for the HL χ^2 test were not supplied.

The (Sepsis-Related) Sequential Organ Failure Assessment Score (SOFA)

SOFA score was developed in 1996 to standardise the assessment of a patient's progress on the ICU during a septic episode.[32] Designed by an expert committee, it grades the dysfunction of each organ system depending on the most abnormal value recorded for parameters chosen to represent those systems. Daily scores for each organ system can be compared separately with previous values or combined into a total score to reflect the overall patient progress.

In 2003 a team from Italy calculated SOFA scores for the first 10 postoperative days in cardiac surgery patients who stayed more than 96 hours in ICU. [33] The worst daily SOFA score, total maximum SOFA and the difference between these two values and the first day SOFA score were calculated. All four derivatives of the SOFA score demonstrated good discrimination with the worst daily score demonstrating the best performance. In 2006 Patila et al. prospectively calculated the SOFA score for 857 cardiac surgery patients.[34] The maximum SOFA score during the first 3 days demonstrated acceptable discrimination for mortality with the overall maximum postoperative SOFA performing slightly better. A 2007 study analysed the association between the day 1 SOFA score and hospital mortality for 1458 cardiac surgery patients and found that the score had acceptable discrimination.[35]

SOFA scores calculated on each of the first six postoperative days, as well as mean and maximum SOFA scores showed good calibration and discrimination in Doerr's study.[23] In a subsequent analysis of the same data, predictions for 30 day mortality made using daily SOFA scores, the maximum SOFA score and the mean of all SOFA scores recorded throughout ICU admission were compared with predictions made using the mean of all daily SOFA scores up to that point. [16] Daily SOFA scores and their derivatives all demonstrated good discrimination.

In Exarchopoulos' validation study, the SOFA score demonstrated acceptable discrimination and calibration when predicting 30 day mortality.[24] Tsaousi et al. studied the accuracy of in-hospital mortality predictions made using day one SOFA scores, maximum and mean SOFA scores and the difference between maximum SOFA and the daily SOFA score. Day one SOFA demonstrated good discrimination but was outperformed by the other SOFA derivatives.[25]

Logistic Organ Dysfunction Score (LODS)

The LODS was developed by Le Gall et al. in 1996.[36] It aimed to predict hospital mortality using a subset of the same database used to develop the SAPS-II score. The LODS uses the worst values recorded during the first 24 hours of ICU admission for 12 variables. Cardiac surgery patients were again excluded. In 2011 Heldwein et al. showed that daily LODS scores could be used to predict mortality in cardiac surgery patients,[37] with the best discrimination observed on the third postoperative day.

Simplified Acute Physiology Score 3 (SAPS-3)

The SAPS-3 score was developed using data from 21336 patients from 309 ICUs across 35 countries[38,39] including 1657 cardiac surgery patients. Variables were selected using a combination of expert opinion and regression modelling. They included existing measures for the classification of illness and physiological instability measured within the first hour of ICU admission. The model is formed of 20 variables, including those reflecting the geographical location of the institution in which it is being used. The total SAPS-3 score is reduced by 6 points for cardiac surgery patients to reflect the greater use of vasoactive drugs and the frequency of abnormal postoperative physiology in these patients. In 2014 Doerr et al. compared SAPS-3 with SAPS-II in 5207 cardiac surgery patients.[29] They calculated the scores on the first six postoperative days and found that SAPS-3 outperformed SAPS-II but was not adequately calibrated when predicting ICU mortality.

Intensive Care National Audit and Research Centre model (ICNARC)

In 2007 Harrison et al. published the ICNARC model,[40] developed using data from 216,626 patients admitted to 163 general ICUs in the UK between 1995 and 2003. The score includes the worst values for 12 variables, six of which were physiological. Cardiac surgery patients were included in the development cohort. In 2015 Ariyaratnam et al. validated the ICNARC model on 1646 cardiac surgery patients in a UK centre and found that it performed well in terms of discrimination and calibration.[3]

Models designed specifically for cardiac surgery

Intensive Care Unit Risk Stratification Score (ICURS)

In 1997 Higgins et al. produced the ICURS based on pre-, intra- and postoperative data recorded on admission to ICU after cardiac surgery for 2440 patients.[17] Separate logistic regression models to predict in-hospital mortality and composite morbidity (defined in terms of specific measures of organ dysfunction) were developed. Eight variables were included in the mortality model and 13 in the morbidity model.

ICURS discriminated well in prospective validation sets, and calibration was reported as good. In 2005 Serrano validated ICURS' ability to predict the duration of mechanical ventilation. ICURS performed best when predicting ventilation lasting more than 48 hours, but discrimination was below the acceptable threshold.[41] In 2006 Biagioli et al. studied the predictions generated by an ICURS model developed using Higgin's methods in their own development cohort. In a separate validation group of 350 cardiac surgery patients this customised model performed poorly.[42] In 2007 Palomba et al. used the ICURS scores of 603 cardiac surgery patients to predict the development of mild AKI with acceptable discrimination.[8]

Cardiac Surgery Score (CASUS)

The Cardiac Surgery Score (CASUS) was developed by Hekmat et al. in 2005 to produce daily 30 day mortality estimates for cardiac surgery patients.[22] The development dataset included 384 patients who underwent cardiac surgery requiring cardiopulmonary bypass followed by admission for >24 hours to ICU. The model based predictions on the most abnormal daily values of 10 variables.

The score was validated in two groups of 1057 and 1104 patients and performed consistently well. In 2010, a subsequent validation using data from 3801 patients, which included the 1104 from the 2005 paper, revealed good discrimination and calibration. CASUS performed best on day 1 and worst on day 5.

Daily CASUS scores, together with mean and maximum CASUS scores, were validated for 30 day mortality prediction at a different German centre in 2011 and were found to perform consistently well over the first six postoperative days.[23] Maximum and mean CASUS scores demonstrated superior discrimination and satisfactory calibration. The same data was used to show that CASUS outperformed SOFA in ICU mortality prediction.[43] The same year a further comparison of CASUS with the new logistic CASUS based on 4054 patients (including the 2801 previously analysed in other studies) was performed.[44] Although discrimination was good, calibration was found to be poor. CASUS was validated in the Exarchopoulos study and demonstrated good discrimination and calibration on the first postoperative day.[24] Log-CASUS[44] and Rapid Clinical Evaluation (RACE)[45], both based on CASUS, performed well in development sets but are yet to be validated themselves.

Biagioli Model

In 2006 Biagioli et al. produced a risk model for cardiac surgery using a Bayes linear approach.[42] The authors trained their model to predict morbidity using data for a range of predictor variables taken from a group of 740 patients undergoing CABG surgery. The final model included pre- and intraoperative data combined with white cell count and oxygen delivery index measured within 3 hours of ICU admission. In a validation set of 350 patients, the model had good discrimination and calibration and outperformed models created using logistic regression.[42]

Salamonsen Model

In 2008 Salamonsen et al. produced a risk model designed to predict which patients undergoing CABG would not be ready for discharge from ICU within their "fast-track" schedule (<12 hours).[46] Pre-, intra- and postoperative variables were used to develop a multiple linear regression model to predict length of stay on the ICU. The model was validated in 117 patients. The R^2 value for the validation set was poor and the 95% confidence intervals for predicted lengths of stay of 4 and 12 hours spanned 29 and 70 hours respectively. Consequently, the authors concluded that their model was not useful.

Meyfroidt Model

In 2011 Meyfroidt *et al.* collected a range of admission, medication, laboratory and physiological data from the first 4 hours of ICU admission for 461 cardiac surgery patients. They used this data to train Gaussian process models to perform two tasks:[47] (i) a classification task to predict whether patients would be discharged from ICU on day 2, (ii), a regression task designed to predict the actual day of ICU discharge. Data for five physiological variables were averaged across 40 minute segments and these averaged values were included in the final model. The models were tested on a validation cohort of 499 patients and were able to adequately identify patients likely to be discharged on day 2 but were less successful when predicting the day of discharge.

Acute Kidney Injury after Cardiac Surgery (AKICS) model

In 2007 Palomba et al. developed and validated a model to predict mild AKI in patients following cardiac surgery.[8] The model was based on eight variables, two of which were postoperative physiological variables. It performed well when validated in 215 patients.

DISCUSSION

This systematic review has identified 14 validated risk models that utilise postoperative patient monitoring data to predict outcomes after adult cardiac surgery. The most commonly validated predictions were for mortality, but the prediction of composite morbidity, ICU length of stay, and specific morbidity outcomes have also been tested. Of the fourteen models, eight were developed on non-cardiac surgery patients but have subsequently been validated in cardiac surgery and six were developed specifically for cardiac surgery.

Postoperative risk prediction models may be useful for performing three main tasks after cardiac surgery. The first is resource allocation where future operating lists and staffing levels may be adjusted according to the predicted length of stay or mortality rates (used as a surrogate for severity of illness) of patients present on the ICU. Secondly, for benchmarking institutional performance where risk estimates can be used to generate standardised predictions for mortality rates against which observed outcomes can be measured. Finally, with caution, risk models may be used to inform clinical decision making and discussions with patients and their relatives. The models identified estimate the risk of adverse outcomes for groups of patients with similar scores. They state the proportion of a group of patients with similar risk scores that would be expected to suffer the outcome. This information may provide a context to clinical decision making and prognostic discussions. Moreover, changes in the predicted risk over time or in response to treatment may give an indication of a patient's progress. However, it should be acknowledged that the scores cannot identify whether or not an individual patient will suffer the outcome.

The majority of models with good discrimination and calibration identified in this review are those which calculate 30-day mortality risk daily based upon the worst value for each parameter in each 24 hour period. While models which predict mortality are potentially useful for benchmarking and resource allocation they are of limited use in guiding real-time treatment decisions. The prediction of specific complications or patient deterioration after cardiac surgery would be much more relevant to the treating clinicians. Such an approach would allow targeted treatment to prevent or reduce the impact of these developing complications.[38] Our review identified only the AKICS score as being capable of predicting acute kidney injury while APACHE-II and SAPS-II successfully predicted prolonged ventilation.

Secondly, these scores are calculated retrospectively once the worst values in a 24 hour period are known; by the time increased risk is detected the complication may be established.[12, 16, 21, 33, 39] Derivative scores such as the mean or maximum value for validated scores over a number of days show even better predictive power.[12, 16, 30] However, due to their retrospective nature these scores also have little value in the day to day treatment of patients. Importantly, serial scores and their aforementioned derivatives are not independent of the quality of care provided by the ICU; poor care will lead to poor mean and maximum scores. They should not be used to produce mortality predictions against which observed mortality rates are measured when benchmarking ICU performance.

A number of models provide a snapshot of risk using data obtained within the first four hours of ICU admission following cardiac surgery. [13, 19, 34, 38, 42, 43] This may be the most appropriate time to estimate risk for the purposes of benchmarking ICU performance. However, these models cannot reliably guide resource allocation or clinical decision making after the initial period on ICU as their predictions may become inaccurate as postoperative events unfold. Some authors validated these models as daily assessment tools to be calculated using the worst scores for each 24 hours with acceptable statistical performance.[16, 20, 21, 33, 40] While statistical performance may be good, as with scores designed for serial use, the predictions are obtained too late to influence patient management and the effect of the quality of ICU care on the scores themselves precludes their use for ICU benchmarking.

Models that would be of most benefit in clinical decision making would utilise up to date clinical information and provide continuously updated predictions, however none of the models identified utilises real-time patient monitoring data. The majority of identified models require the most abnormal value for each parameter over a given period and categorise continuous variables according to the degree of abnormality. This approach sacrifices predictive accuracy to improve the ease of use and minimise the need for computing power.[13, 38, 40] With recent developments in computing more ambitious approaches may be possible. The model developed by Meyfroidt utilising Gaussian processes does use computerised analyses of a large number of data points.[43] However, even this model analysed average values calculated for 40 minute periods rather than continuous data.

This review has demonstrated that models developed for use in general ICU patients such as the SOFA, SAPS-II and APACHE-II scores may be applied successfully to cardiac surgery patients.[22-24, 31, 34]. This is despite their developers' excluding cardiac surgery patients from development datasets due to their low observed mortality when compared with other patient groups with similar levels of physiological derangement.[19]

However, there may be advantages to using cardiac surgery specific scores. Firstly, a number of risk factors included in general ICU models such as metastatic cancer and liver cirrhosis are largely irrelevant, as they usually contraindicate cardiac surgery. In addition, there are many significant differences in care protocols between cardiac and general ICU's. For example, the CASUS developers noted that the conscious level of patients is routinely decreased in the early postoperative period secondary to sedation. Therefore, they introduced a 'neurological state score' which was quicker and easier to calculate than the Glasgow Coma Scale and decreased the impact of appropriately low conscious level on risk estimates. They also recognised the need to correct risk scores for artificially normal physiological values which are only present as a direct consequence of supportive treatments frequently used following cardiac surgery, such as mechanical cardiovascular support or renal replacement therapy.[22] As a result, despite general ICU models demonstrating good statistical performance, a cardiac surgery specific model may be preferred by clinicians.

There are however, a number of key limitations of the cardiac surgery specific models identified in this review which are likely to explain their limited adoption. First, unlike widely used pre-operative cardiac surgery risk models[1, 2] and the models developed for general ICUs, most cardiac surgery models have been based on data from single centres. This approach optimises data quality and completeness for model development but may lead to concerns about the application of models to different populations. For example, the Biagioli, ICURS, Meyfroidt and AKICS models require cardiac output measurement using a Swann-Ganz catheter which is not routinely used in all cardiac surgery centres.[48] The Meyfroidt model also contains variables derived from entropy measurements. These values describe the variation within a patient's physiological data, but monitoring equipment capable of producing these values may not be available in all ICUs. Similarly, when initiation of specific treatments e.g. intra-aortic balloon counterpulsation, are used as surrogates for severity of physiological derangement, local practices can affect the validity of these surrogate variables. The cardiovascular component of the SOFA score is based on the administration of vasoactive medication using specific protocols (such as dopamine being administered before noradrenaline to treat hypotension). In many centres clinicians will know that these patterns of drug administration are not followed and this may lead to diminished confidence in the SOFA score despite reports of good performance in multiple studies.[16,23-25,34,35,43]

CONCLUSION

Risk prediction models based on preoperative data have real value when advising patients on their decision whether to undergo surgery and when assessing the performance of cardiothoracic units. However, postoperative models identified in this review have the key advantage of being updated throughout a patient's admission. If they are used to produce risk estimates at the time of admission to ICU, they may be used to assess the quality of the ICU care in isolation from the pre- and intraoperative events. Models which produce daily risk estimates deliver updated predictions which enable optimisation of resource allocation planning in cardiac surgery units. As described in this review, most of the models make predictions which are accurate enough to perform these two tasks. SOFA and CASUS are the most extensively validated scores and use readily available postoperative variables to produce their risk estimates. This combination of ease of calculation and accuracy defines them as the most appropriate postoperative scores identified in this study. Their discriminatory power is beyond that displayed by preoperative scores such as EuroSCORE and EuroSCORE II. [4,49] With caution, these scores may also be used to inform discussions with patients and their relatives and provide a broad context for clinical decision making.

However, no existing model provides estimates for the risk of specific complications for individuals with sufficient accuracy and frequency to reliably guide specific clinical decisions. This is probably the main reason why such models have not achieved widespread adoption into clinical practice.

Technological developments have the potential to improve risk prediction after cardiac surgery. In future, computerised models designed to calculate risk much more frequently could provide contemporaneous risk estimates. The most useful models would predict specific complications early enough to allow clinicians time to intervene to prevent the complications occurring or, where that is not possible, reduce their impact. The ideal model would analyse physiological variables and not the current treatments, thus avoiding the pitfall of interinstitutional variation in management protocols. Variables could be selected from the huge amount of post-cardiac surgery data available on the ICU according to the specific outcome being predicted. The accuracy of such models may be improved by advances in computing which enable real-time analysis of raw monitoring data rather than categorical “worst values" recorded over a given time period. Analyses of changes in, rather than absolute values of, an individual’s physiological variables may allow identification of those at increased risk of clinical deterioration before arbitrary thresholds for abnormality are reached and end organ damage occurs.

REFERENCES

1. Nashef SA, Roques F, Michel P, et al.: European system for cardiac operative risk evaluation (EuroSCORE). Eur J Cardiothorac Surg 16:9-13, 1999
2. Nashef SA, Roques F, Sharples LD, et al.: EuroSCORE II. Eur J Cardiothorac Surg 41:734-744; discussion 744-735, 2012
3. Ariyaratnam P, Loubani M, Biddulph J, et al.: Validation of the intensive care national audit and research centre scoring system in a UK adult cardiac surgery population. Journal of Cardiothoracic and Vascular Anesthesia 29:565-569, 2015
4. Grant SW, Hickey GL, Dimarakis I, et al.: How does EuroSCORE II perform in UK cardiac surgery; an analysis of 23 740 patients from the Society for Cardiothoracic Surgery in Great Britain and Ireland National Database. Heart (British Cardiac Society) 98:1568-1572, 2012
5. Hosseinian L, Chiang Y, Itagaki S, et al.: Earlier versus later tracheostomy in patients with respiratory failure after cardiac surgery in the United States. J Cardiothorac Vasc Anesth 28:488-492, 2014
6. Bailey ML, Richter SM, Mullany DV, et al.: Risk factors and survival in patients with respiratory failure after cardiac operations. The Annals of thoracic surgery 92:1573-1579, 2011
7. Thakar CV, Worley S, Arrigain S, et al.: Influence of renal dysfunction on mortality after cardiac surgery: Modifying effect of preoperative renal function. Kidney Int 67:1112-1119, 2005
8. Palomba H, de Castro I, Neto AL, et al.: Acute kidney injury prediction following elective cardiac surgery: AKICS Score. Kidney Int 72:624-631, 2007
9. Kogan A, Cohen J, Raanani E, et al.: Readmission to the intensive care unit after "fast-track" cardiac surgery: risk factors and outcomes. Annals of Thoracic Surgery 76:503-507, 2003
10. LaPar DJ, Gillen JR, Crosby IK, et al.: Predictors of operative mortality in cardiac surgical patients with prolonged intensive care unit duration. Journal of the American College of Surgeons 216:1116-1123, 2013
11. Vohra HA, Goldsmith IR, Rosin MD, et al.: The predictors and outcome of recidivism in cardiac ICUs. European Journal of Cardio-Thoracic Surgery 27:508-511, 2005
12. Schottler J, Cremer J, Hagemann A, et al.: Long-term survival after cardiac surgery with complicated postoperative course. Thoracic and cardiovascular surgeon 58,
13. Morgan RJM WF, Wright MM. : An early warning scoring system for detecting developing critical illness. Clin Intensive Care:100, 1997
14. Hanley JA, McNeil BJ: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29-36, 1982
15. Angus DC, Clermont G, Kramer DJ, et al.: Short-term and long-term outcome prediction with the Acute Physiology and Chronic Health Evaluation II system after orthotopic liver transplantation. Critical Care Medicine 28:150-156, 2000
16. Badreldin A, Elsobky S, Lehmann T, et al.: Daily-mean-SOFA, a new derivative to increase accuracy of mortality prediction in cardiac surgical intensive care units. Thoracic and cardiovascular surgeon 60:43-50, 2012
17. Higgins TL, Estafanous FG, Loop FD, et al.: ICU admission score for predicting morbidity and mortality risk after coronary artery bypass grafting. Annals of Thoracic Surgery 64:1050-1058, 1997
18. Hosmer DW, Lemeshow S: Applied logistic regression: Wiley, 1989
19. Knaus WA, Draper EA, Wagner DP, et al.: APACHE II: a severity of disease classification system. Crit Care Med 13:818-829, 1985
20. Knaus WA, Zimmerman JE, Wagner DP, et al.: APACHE-acute physiology and chronic health evaluation: a physiologically based classification system. Crit Care Med 9:591-597, 1981
21. Kern H, Redlich U, Hotz H, et al.: Risk factors for prolonged ventilation after cardiac surgery using APACHE II, SAPS II, and TISS: Comparison of three different models. Intensive care medicine 27:407-415, 2001

22. Hekmat K, Kroener A, Stuetzer H, et al.: Daily assessment of organ dysfunction and survival in intensive care unit cardiac surgical patients. Annals of Thoracic Surgery 79:1555-1562, 2005
23. Doerr F, Badreldin AM, Heldwein MB, et al.: A comparative study of four intensive care outcome prediction models in cardiac surgery patients. Journal of cardiothoracic surgery 6, 2011
24. Exarchopoulos T, Charitidou E, Dedeilias P, et al.: Scoring Systems for Outcome Prediction in a Cardiac Surgical Intensive Care Unit: A Comparative Study. American Journal of Critical Care 24:327-334, 2015
25. Tsaousi GG, Pitsis AA, Ioannidis GD, et al.: Implementation of EuroSCORE II as an adjunct to APACHE II model and SOFA score, for refining the prognostic accuracy in cardiac surgical patients. Journal of Cardiovascular Surgery 56:919-927, 2015
26. Knaus WA, Wagner DP, Draper EA, et al.: THe apache iii prognostic system. risk prediction of hospital mortality for critically ill hospitalized adults. Chest 100:1619-1636, 1991
27. Becker RB, Zimmerman JE, Knaus WA, et al.: The use of APACHE III to evaluate ICU length of stay, resource use, and mortality after coronary artery by-pass surgery. J Cardiovasc Surg (Torino) 36:1-11, 1995
28. Le Gall JR, Lemeshow S, Saulnier F: A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. Jama 270:2957-2963, 1993
29. Doerr F, Badreldin AM, Can F, et al.: SAPS 3 is not superior to SAPS 2 in cardiac surgery patients. Scandinavian cardiovascular journal : SCJ 48:111-119, 2014
30. Marshall JC, Cook DJ, Christou NV, et al.: Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. Crit Care Med 23:1638-1652, 1995
31. Hekmat K, Doerr F, Kroener A, et al.: Prediction of mortality in intensive care unit cardiac surgical patients. European Journal of Cardio-Thoracic Surgery 38:104-109, 2010
32. Vincent JL, Moreno R, Takala J, et al.: The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 22:707-710, 1996
33. Ceriani R, Mazzoni M, Bortone F, et al.: Application of the sequential organ failure assessment score to cardiac surgical patients. Chest 123:1229-1239, 2003
34. Patila T, Kukkonen S, Vento A, et al.: Relation of the Sequential Organ Failure Assessment Score to Morbidity and Mortality After Cardiac Surgery. Annals of Thoracic Surgery 82:2072-2078, 2006
35. Gomes RV, Tura B, Mendonca HTH, et al.: A first postoperative day predictive score of mortality for cardiac surgery. Annals of Thoracic & Cardiovascular Surgery 13:159-164, 2007
36. Le Gall JR, Klar J, Lemeshow S, et al.: The Logistic Organ Dysfunction system. A new way to assess organ dysfunction in the intensive care unit. ICU Scoring Group. Jama 276:802-810, 1996
37. Heldwein MB, Badreldin AM, Doerr F, et al.: Logistic Organ Dysfunction Score (LODS): a reliable postoperative risk management score also in cardiac surgical patients? Journal of cardiothoracic surgery 6, 2011
38. Metnitz PG, Moreno RP, Almeida E, et al.: SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description. Intensive Care Med 31:1336-1344, 2005
39. Moreno RP, Metnitz PG, Almeida E, et al.: SAPS 3--From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. Intensive Care Med 31:1345-1355, 2005
40. Harrison DA, Parry GJ, Carpenter JR, et al.: A new risk prediction model for critical care: The Intensive Care National Audit & Research Centre (ICNARC) model*. Critical Care Medicine 35:1091-1098, 2007
41. Serrano N, Garcia C, Villegas J, et al.: Prolonged intubation rates after coronary artery bypass surgery and ICU risk stratification score. Chest 128:595-601, 2005

42. Biagioli B, Scolletta S, Cevenini G, et al.: A multivariate Bayesian model for assessing morbidity after coronary artery surgery. Critical Care 10, 2006
43. Badreldin AM, Doerr F, Ismail MM, et al.: Comparison between Sequential Organ Failure Assessment score (SOFA) and Cardiac Surgery Score (CASUS) for mortality prediction after cardiac surgery. Thoracic & Cardiovascular Surgeon 60:35-42, 2012
44. Doerr F, Badreldin AM, Bender EM, et al.: Outcome prediction in cardiac surgery: the first logistic scoring model for cardiac surgical intensive care patients. Minerva anestesiologica 78:879-886, 2012
45. Badreldin AM, Doerr F, Bender EM, et al.: Rapid clinical evaluation: an early warning cardiac surgical scoring system for hand-held digital devices. European Journal of Cardio-Thoracic Surgery 44:992-997; discussion 997-998, 2013
46. Salamonsen MR, Bailey MJ, Salamonsen RF: Can ICU stay be predicted accurately enough for fast-tracking cardiac surgical patients? Critical care and resuscitation : journal of the Australasian Academy of Critical Care Medicine 10:194-201, 2008
47. Meyfroidt G, Guiza F, Cottem D, et al.: Computerized prediction of intensive care unit discharge after cardiac surgery: development and validation of a Gaussian processes model. BMC Medical Informatics & Decision Making 11, 2011
48. Bignami E, Belletti A, Moliterni P, et al.: Clinical practice in perioperative monitoring in adult cardiac surgery: is there a standard of care? Results from an national survey. Journal of Clinical Monitoring and Computing:1-19, 2015
49. Chalmers J, Pullan M, Fabri B, et al.: Validation of EuroSCORE II in a modern cohort of patients undergoing cardiac surgery. Eur J Cardiothorac Surg 43:688-694, 2013

Appendix A - Search strategy details

Embase Search

("heart surg*” OR "cardi* surg*” OR coronary adj3 bypass OR "coronary graft” OR “CABG” OR (valv* adj3 (rep* OR surg*)).ti,ab OR *HEART SURGERY/ OR *CORONARY ARTERY BYPASS GRAFT/ OR *MITRAL VALVE REPLACEMENT/ OR *MITRAL ANNULOPLASTY/ OR *HEART TRANSPLANTATION/ OR *VALVULOPLASTY/ OR *CORONARY ARTERY BYPASS SURGERY/

AND

(morbidity OR mortality OR "renal failure" OR "renal replacement" OR "kidney injury" OR arrhythmia OR bleeding OR resternotomy OR "re-sternotomy" OR "respiratory failure" OR fail* adj3 extubation OR fibrillation OR death OR length of stay OR (renal AND replacement AND therapy) OR (prolonged adj3 ventilation) OR fibrillation).ti,ab OR SURGICAL MORTALITY/ OR KIDNEY FAILURE/ OR RENAL REPLACEMENT THERAPY/ OR REOPERATION/ OR POSTOPERATIVE COMPLICATION/ OR HEART TAMPONADE/ OR MORBIDITY/ OR LENGTH OF STAY/ OR DEATH/ OR HEART ARRHYTHMIA/ OR HEART ATRIUM FIBRILLATION/

AND

("intensive care" OR "critical care").ti,ab OR INTENSIVE CARE/

AND

(Predict* OR realtime OR "statistical model" OR "regression model" OR algorithm OR "risk stratification" OR "early identification).ti,ab OR CLINICAL DECISION MAKING/ OR DECISION SUPPORT SYSTEM/ OR MEDICAL DECISION MAKING/ OR COMPUTER SYSTEM/ OR PREDICTION AND FORECASTING/ OR *RISK ASSESSMENT/

Medline Search

("heart surg*" OR "cardi* surg*" OR "coronary artery bypass" OR "coronary bypass" OR "coronary graft" OR “CABG” OR (valv* adj3 (replac* OR repair OR surg*)).ti,ab OR exp *CARDIAC VALVE ANNULOPLASTY/ OR exp *CORONARY ARTERY BYPASS/ OR *CARDIAC SURGICAL PROCEDURES/ OR *HEART TRANSPLANTATION/ OR *HEART VALVE PROSTHESIS/

AND

(morbidity OR mortality OR("renal failure" OR "renal replacement" OR arrhythmia* OR bleeding OR resternotomy OR "re-sternotomy" OR "respiratory failure" OR fail* adj3 extubation OR death.ti,ab OR "kidney injury" OR prolonged adj3 ventilation OR fibrillation OR (failed AND extubation) OR "length of stay".ti,ab OR MORBIDITY/ OR MORTALITY/ OR HOSPITAL MORTALITY/ OR RENAL INSUFFICIENCY/ OR *ACUTE KIDNEY INJURY/ OR *RENAL REPLACEMENT THERAPY/ OR *REOPERATION/ OR *POSTOPERATIVE COMPLICATIONS/ OR *CARDIAC TAMPONADE/ OR *RESPIRATORY INSUFFICIENCY/ OR *DEATH OR *ARRHYTHMIAS, CARDIAC/ OR *ATRIAL FIBRILLATION/ OR *ATRIAL FLUTTER/ OR RENAL REPLACEMENT THERAPY/ OR RENAL DIALYSIS/ or HEMOFILTRATION/ or TRACHEOSTOMY/ OR LENGTH OF STAY/

AND

("intensive care OR "critical care")ti,ab OR *CRITICAL CARE/ OR *INTENSIVE CARE/

AND

(Predict* OR realtime OR "statistical model" OR "regression model" OR algorithm OR "risk stratification" OR "early identification").ti,ab OR *DECISION MAKING, COMPUTER-ASSISTED/ OR *DECISION SUPPORT SYSTEMS, CLINICAL/ OR *COMPUTER SYSTEMS/

Cochrane Library Search

“Cardi* surg*” OR CABG OR “Coronary Artery Bypass” OR “Heart surg*” OR "coronary graft" OR “Coronary bypass” OR Valv* adj3 (replac* or repair or surg*) OR (MeSH descriptor: [Coronary Artery Bypass] explode all trees) OR (MeSH descriptor: [Thoracic Surgery] explode all trees) OR (MeSH descriptor: [Cardiac Valve Annuloplasty] explode all trees) OR (MeSH descriptor: [Heart Valve Prosthesis Implantation] explode all trees) OR (MeSH descriptor: [Cardiac Surgical Procedures] explode all trees)

AND

morbidity OR mortality OR "renal failure" OR "renal replacement" OR arrythmia OR bleeding OR resternotomy OR "re-sternotomy" OR "respiratory failure" OR fail* adj extubation OR "kidney injury" OR death OR "length of stay" OR prolonged adj3 ventilation OR (MeSH descriptor: [Morbidity] explode all trees) OR (MeSH descriptor: [Mortality] explode all trees)OR (MeSH descriptor: [Renal Insufficiency] explode all trees) OR (MeSH descriptor: [Acute Kidney Injury] explode all trees)OR (MeSH descriptor: [Renal Replacement Therapy] explode all trees)OR (MeSH descriptor: [Postoperative Complications] explode all trees)OR (MeSH descriptor: [Reoperation] explode all trees)OR (MeSH descriptor: [Respiratory Insufficiency] explode all trees) OR (MeSH descriptor: [Cardiac Tamponade] explode all trees) OR (MeSH descriptor: [Death] explode all trees) OR (MeSH descriptor: [Arrhythmias, Cardiac] explode all trees) OR(MeSH descriptor: [Tracheostomy] explode all trees) OR (MeSH descriptor: [Length of Stay] explode all trees)

AND

realtime OR "statistical model" OR "regression model" OR algorithm OR "risk prediction" OR "risk stratification" OR "early identification" OR (MeSH descriptor: [Decision Making, Computer-Assisted] explode all trees) OR (MeSH descriptor: [Decision Support Techniques] explode all trees)

AND

"intensive care" OR "critical care" OR (MeSH descriptor: [Critical Care] explode all trees)

Table 1

Table 1: Models validated for predicting outcomes following cardiac surgery

Model	Author	Year	Country	Development method	Design cohort	Cardiac surgery validation	Outcomes predicted	No. of physiological parameters
APACHE-II[19]	Knaus	1985	USA	Logistic regression	Excluded cardiac	External	Perioperative, ICU and 30 day Mortality; LOS-ICU ; Prolonged mechanical ventilation	5
APACHE-III[27]	Knaus	1991	USA	Logistic regression	Excluded cardiac	External	Hospital mortality; LOS-ICU; Treatment costs	5
SAPS-II[28]	Le Gall	1993	12 countries	Logistic regression	Excluded cardiac	External	Hospital and ICU Mortality; Prolonged mechanical ventilation	3
MODS[30]	Marshall	1995	Canada	Logistic regression	Surgical ICU	External	Mortality	5
SOFA[32]	Vincent	1996	16 countries	Expert Opinion	General ICU	External	Hospital and ICU Mortality; LOS-ICU	3
LODS[36]	Le Gall	1996	12 countries	Logistic regression	Excluded cardiac	External	Hospital and ICU mortality	5
ICURS[17]	Higgins	1997	USA	Logistic regression	Mixed cardiac	External	Hospital Mortality; Composite morbidity	4
SAPS-3[38]	Moreno	2005	35 countries	Logistic regression	General ICU	External	Hospital and ICU mortality	2
CASUS[22]	Hekmat	2005	Germany	Logistic regression	Mixed cardiac	Internal/ External	30 day and ICU mortality	5
Biagioli[42]	Biagioli	2006	Italy	Bayesian	CABG	Internal	Composite morbidity	2
ICNARC[40]	Harrison	2007	UK	Logistic regression	General ICU	External	Perioperative mortality	7
Salamonsen[46]	Salamonsen	2008	Australia	Linear regression	CABG	Internal	LOS-ICU	3
Meyfroidt[47]	Meyfroidt	2011	Belgium	Gaussian process	Mixed cardiac	Internal	LOS-ICU	13*
AKICS[8]	Palomba	2007	Brazil	Logistic regression	Mixed cardiac	Internal	AKI	2

APACHE-II – Acute Physiology and Chronic Health Evaluation-II, APACHE-III – Acute Physiology and Chronic Health Evaluation-III, SAPS-II – Simplified Acute Physiology Score II, MODS – Multiple Organ Dysfunction Score , SOFA – (Sepsis-Related) Sequential Organ Failure Assessment, LODS – Logistic Organ Dysfunction Score, ICURS – Intensive Care Unit Risk Stratification Score, SAPS-3 – Simplified Acute Physiology Score 3, CASUS – Cardiac Surgery Score, ICNARC – Intensive Care National Audit and Research Centre, ICU Intensive Care Unit, AKICS – Acute Kidney Injury after Cardiac Surgery, CABG – Coronary Artery Bypass Graft, LOS-ICU – Length Of Stay on the Intensive Care Unit, AKI – Acute Kidney Injury

* Included multiple statistical values for parameters including means, variances and cumulative totals

Table 2

Table 2 -Variables included in each model

Model	Pre-operative	Intra-operative	Postoperative physiological	Other Postoperative	Timing of capture
APACHE-II[19]	Age, Chronic Disease Status, type of admission	-	PaO_2/FiO_2, Temp, MAP, RR	Blood tests: pH, WCC, K^+, Na^+, Hct, Cr GCS, FiO_2	Worst value recorded each day (originally within first 24hours)
APACHE-III[27]	Age, Previous surgery, Gender, Comorbidities	Number of grafts and vessels used. Urgency	HR, MAP, Temp, RR , A-a gradient, UO	Blood tests: Hct, WCC, Cr, Na^+, Albumin, Bilirubin, glucose, BUN, PaO_2	Worst value recorded within first 24hours
SAPS-II[28]	Age, Chronic Disease Status, Type of Admission	-	PaO_2/FiO_2, UO	Blood tests: Ur, Cr, WCC, K^+, Na^+, HCO_3^- GCS	Worst value recorded each day (originally within first 24hours)
MODS[30]	-	-	PaO_2/FiO_2, PAR	Blood tests: Bilirubin, Cr, Platelets GCS	Worst value recorded each day
SOFA[32]	-	-	PaO_2/FiO_2, MAP	Blood Tests: Cr, Bilirubin, Platelets, Vasopressor use, GCS	Worst value recorded each day
LODS[36]	-	-	PaO_2/FiO_2, HR, systolic BP, UO	Blood tests: WCC, Ur, Cr, Bilirubin, PT, Platelets, GCS	Worst value recorded each day (originally within first 24hours)
ICURS[17]	Age, Comorbidities, Albumin	CPB time Need for IABP after CPB	A-a gradient, HR, CI,	Blood tests: HCO_3^-	On arrival to ICU
SAPS-3[38]	Age, Comorbidities, Reason for Admission, Pre-admission events	Site of surgery	Temp, HR	Blood tests: Bilirubin, Cr, WCC, pH, Platelets GCS, FiO2, requirement for mechanical ventilation	Within 1 hour of admission
CASUS[22]	-	-	PaO_2/FiO_2, PAR	Blood tests: Cr, Bilirubin, lactate, Platelets. Neurological state, Requirement for IABP or VAD	Worst value recorded each day

Biagioli[42]	Age, Weight, Comorbidities, Cr, Requirement for IABP	Type of surgery Duration of CPB	DO_2I,	Blood tests: WCC Requirement for IABP	Within 3 hours of admission
ICNARC[40]	Age, diagnostic category, source of admission, CPR before admission	-	HR, systolic BP, Temp, RR, PaO_2/FiO_2, UO,	pH, Ur, Cr, Na, WCC, GCS	Within 24 hours of admission
Salamonsen[46]	-	-	MAP, CVP, CI	Blood tests: HCO_3^- Requirement for IABP Cumulative adrenaline and noradrenaline doses	Average values over first four hours on ICU
Meyfroidt[47] *	Comorbidities, Pre-admission events	-	Multiple derived from BP, RR, FiO_2, SpO_2, PAP, PEEP, HR, CVP, SPAP, UO, Drain Output, CO, Temp	Blood tests Medication	First four hours of admission
AKICS[8]	Age, Cr, Glucose, type of surgery, comorbidities	Duration of CPB	CO, CVP		On ICU admission

*see http://www.kuleuven.be/licm/ml/gpdischarge1.html for details of modelled variables

APACHE-II – Acute Physiology and Chronic Health Evaluation-II, APACHE-III – Acute Physiology and Chronic Health Evaluation-III, SAPS-II – Simplified Acute Physiology Score II, MODS – Multiple Organ Dysfunction Score , SOFA – (Sepsis-Related) Sequential Organ Failure Assessment, LODS – Logistic Organ Dysfunction Score, ICURS – Intensive Care Unit Risk Stratification Score, SAPS-3 – Simplified Acute Physiology Score 3, CASUS – Cardiac Surgery Score, ICNARC – Intensive Care National Audit and Research Centre, ICU Intensive Care Unit, AKICS – Acute Kidney Injury after Cardiac Surgery

A-a gradient – alveolar arterial gradient, Albumin – serum albumin concentration, Bilirubin - serum bilirubin concentration, BP – blood pressure, BUN – blood urea nitrogen, CI – Cardiac index, CO – cardiac output, CPB – cardiopulmonary bypass, Cr – serum creatinine concentration, CVP – central venous pressure, DO_2I – oxygen delivery index, FiO2 – fraction inspired oxygen, GCS – Glasgow Coma Scale, Glucose – serum glucose concentration, Hct – haematocrit, $HCO3^-$ – serum bicarbonate concentration, IABP – Intra-aortic balloon pump, K^+ – serum potassium concentration, lactate – serum lactate concentration, MAP – mean arterial pressure, Na^+ – serum sodium concentration, PAR – pressure adjusted heart rate (HRxCVP/MAP), PaCO2 – arterial partial pressure of carbon dioxide, PaO2 – arterial partial pressure of oxygen, PAP – Peak airway pressure, PCWP – pulmonary capillary wedge pressure, PEEP – positive end-expiratory pressure, pH – blood pH, Platelets – Platelet Count, RR – respiratory rate, SPAP – systolic pulmonary artery pressure, Temp – temperature, UO – urine output, Ur – serum urea concentration, VAD – ventricular assist device, VT – Tidal Volume, WCC – White Cell Count

Table 3

Table 3 - Studies validating models in the prediction of mortality in cardiac surgery

Model	Author	Year	Country	Validation cohort (n)	Measure of calibration*	Measure of discrimination*
APACHE-II	Hekmat[22]^	2005	Germany	Mixed cardiac (1057)	HL χ^2=6.6‡	AUC=0.89
	Doerr[23]	2011	Germany	Mixed cardiac (2801)	HL χ^2=30.6 (p<0.001)	AUC=0.87
	Ariyaratnam[3]	2015	UK	Mixed cardiac (1646)	HL χ^2=16.2 (p=0.001)	AUC=0.65
	Exarchopoulos[24]	2015	Greece	Mixed cardiac (150)	HL χ^2=10.9 (p=0.20)	AUC=0.82
	Tsaousi[25]	2015	Greece	Mixed cardiac (1058)	HL χ^2=7.4 (p=0.49)	AUC=0.86
APACHE-III	Becker[27]	1995	USA	Mixed cardiac (2435)	R^2=0.22	AUC=0.85
SAPS-II	Doerr[23]	2011	Germany	Mixed cardiac (2801)	HL χ^2=17.15 (p=0.03)	AUC=0.89
	Doerr[29]	2014	Germany	Mixed cardiac (5207)	HL χ^2=57.8 (p=0.000)	AUC=0.88
	Exarchopoulos[24]	2015	Greece	Mixed cardiac (150)	HL χ^2=5.1 (p=0.75)	AUC=0.80
MODS	Hekmat[22]^	2005	Germany	Mixed cardiac (1057)	HL χ^2=5.7‡	AUC=0.90
SOFA	Ceriani[33]	2003	Italy	Mixed cardiac (218)		AUC=0.71
	Patila[34]#	2006	Finland	Mixed cardiac (855)		AUC=0.78
	Gomes[35]	2007	Brazil	Mixed cardiac (1458)		AUC=0.74
	Doerr[23]	2011	Germany	Mixed cardiac (2801)	HL χ^2=6.75 (p=0.56)	AUC=0.91
	Badreldin[43]	2012	Germany	Mixed cardiac (2801)	HL χ^2=6.75 (p=0.56)	AUC=0.91
	Badreldin[16]	2012	Germany	Mixed cardiac (2801)	HL χ^2=14.9 (p=0.06)	AUC=0.88
	Exarchopoulos[24]	2015	Greece	Mixed cardiac (150)	HL χ^2=2.9 (p=0.57)	AUC=0.76
	Tsaousi[25]	2015	Greece	Mixed cardiac (1058)	HL χ^2=4.8 (p=0.58)	AUC=0.86
LODS	Heldwein[37]	2011	Germany	Mixed cardiac (2801)	HL χ^2=6.4 (p=0.49)	AUC=0.93
ICURS	Higgins[17]	1997	USA	Mixed cardiac (2125)	Good HL χ^2‡	AUC=0.85
	Gomes[35]	2007	Brazil	Mixed cardiac (1458)		AUC=0.77
SAPS-3	Doerr[29]	2014	Germany	Mixed cardiac (5207)	HL χ^2=15.2 (p=0.056)	AUC=0.89
CASUS	Hekmat[22]^	2005	Germany	Mixed cardiac (1104)	HL χ^2 =5.1‡	AUC=0.96
	Hekmat[31]^	2010	Germany	Mixed cardiac (3801)	HL χ^2 =7.0‡	AUC=0.95
	Doerr[23]	2011	Germany	Mixed cardiac (2801)	HL χ^2=14.0 (p=0.05)	AUC=0.97
	Badreldin[43]	2012	Germany	Mixed cardiac (2801)	HL χ^2=14.0 (p=0.05)	AUC=0.97
	Doerr[44]	2012	Germany	Mixed cardiac (4054)	O/E ratio=0.63	AUC=0.97
	Exarchopoulos[24]	2015	Greece	Mixed cardiac (150)	HL χ^2=2.2 (p=0.89)	AUC=0.89
ICNARC	Ariyaratnam[3]	2015	UK	Mixed cardiac (1646)	HL χ2 =9.10 (p=0.33)	AUC=0.85

APACHE-II – Acute Physiology and Chronic Health Evaluation-II, APACHE-III – Acute Physiology and Chronic Health Evaluation-III, SAPS-II – Simplified Acute Physiology Score II, MODS – Multiple Organ Dysfunction Score , SOFA – (Sepsis-Related) Sequential Organ Failure Assessment, LODS – Logistic Organ Dysfunction Score, ICURS – Intensive Care Unit Risk Stratification Score, SAPS-3 – Simplified Acute Physiology Score 3, CASUS – Cardiac Surgery Score, ICNARC – Intensive Care National Audit and Research Centre, ICU Intensive Care Unit, AKICS – Acute Kidney Injury after Cardiac Surgery
HL – Hosmer Lemeshow, AUC - Area under the receiver operating characteristic curve
*if calculated on multiple days the value on the day of the best AUC is shown

‡ p values not supplied
# only investigated maximum SOFA score
^ if multiple similar samples of patients were studied in the same paper the values for the biggest sample are shown

Table 4

Table 4 - Studies validating models in the prediction of morbidity in cardiac surgery

Model	Author	Year	Country	Validation cohort (n)	Measure of calibration*	Measure of discrimination*
Length of ICU stay						
APACHE-III	Becker[27]	1995	USA	Mixed Cardiac (2435)	R^2=0.08	
Salamonsen	Salamonsen[46]	2008	Australia	CABG (117)	R^2=0.38	
Meyfroidt	Meyfroidt[47]	2011	Belgium	Mixed cardiac (499)	HL χ2 good (p=0.38) Brier 0.18	AUC=0.76
Composite morbidity						
ICURS	Higgins[17]	1997	USA	Mixed cardiac (2125)	Good HL χ^2‡	AUC=0.76
	Biagioli[42]	2006	Italy	CABG (740)	Poor HL χ^2‡	AUC=0.82
Biagioli	Biagioli[42]	2006	Italy	CABG (350)	HLχ^2 good (p=0.35)	AUC=0.70
Acute Kidney Injury						
ICURS	Palomba[8]	2007	Brazil	Mixed cardiac (603)		AUC=0.70
AKICS	Palomba[8]	2007	Brazil	Mixed cardiac (215)	HL χ2 good (p=0.24)	AUC=0.79
Prolonged Mechanical Ventilation						
SAPS-II	Kern[21]	2001	Germany	Mixed cardiac (687)		AUC=0.90
APACHE-II	Kern[21]	2001	Germany	Mixed cardiac (687)		AUC=0.88
ICURS	Serrano[41]	2005	Spain	CABG (569)	HL χ^2= 12.1 (p=0.10)	AUC=0.68

APACHE-II – Acute Physiology and Chronic Health Evaluation-II, APACHE-III – Acute Physiology and Chronic Health Evaluation-III, SAPS-II – Simplified Acute Physiology Score II, ICURS – Intensive Care Unit Risk Stratification Score, ICU Intensive Care Unit, AKICS – Acute Kidney Injury after Cardiac Surgery

CABG, Coronary Artery Bypass Graft, HL – Hosmer Lemeshow, AUC - Area under the receiver operating characteristic curve
*if calculated on multiple days the value on the day of the best AUC is shown
‡ p values not supplied

Table 5

Table 5 –Validation studies-quality

Validation study and Year	Models Validated	Patient selection criteria detailed	Consecutive Patients Studied	Preop health status well described	Patient demographics well described	Data collection	Handling of missing data	Outcome measures	Validation method	Validation group size
Higgins 1997[17]	ICURS	Yes	Yes	Yes	Yes	Prospective	Not discussed	Mortality and Composite Morbidity	Internal	2125
Kern 2001[21]	SAPS-II, APACHE-II	Yes	Yes	No	Yes	Prospective	Not discussed	Prolonged Mechanical Ventilation	External	687
Ceriani 2003[33]	SOFA	Yes	Yes	Yes	Yes	Not specified	Not discussed	Mortality	External	218
Serrano 2005[41]	ICURS	Yes	Yes	No	Yes	Prospective	Not discussed	Prolonged Mechanical Ventilation	External	569
Hekmat 2005[22]	APACHE-II,MODS	Yes	Yes	No	Yes	Prospective	No missing data	Mortality	External	1057
	CASUS	Yes	Yes	No	Yes	Prospective	No missing data	Mortality	Internal	1057
Patila 2006[34]	SOFA	Yes	Yes	Yes	Yes	Prospective	Not discussed	Mortality	External	857
Biagioli 2006[42]	locally customised ICURS	Yes	Yes	Yes	Yes	Prospective	Not discussed	Composite Morbidity	Internal	350
	Biagioli	Yes	Yes	Yes	Yes	Prospective	Not discussed	Composite Morbidity	Internal	350
Gomes 2007[35]	SOFA	Yes	Yes	Yes	Yes	Not specified	Not discussed	Mortality	External	1458
Palomba 2007[8]	ICURS	Yes	Not specified	Yes	Yes	Prospective	Not discussed	AKI	External	603
	AKICS	Yes	Not specified	Yes	Yes	Prospective	Not discussed	AKI	Internal	215
Salamonsen 2008[46]	Salamonsen	Yes	Not specified	Yes	Yes	Prospective	Patients excluded	LOS-ICU	Internal	117

Hekmat 2010[31]	CASUS	Yes	Yes	No	Yes	Prospective	No missing data	Mortality	Internal	3801
Doerr 2011[23]	CASUS, SOFA, SAPS-II APACHE-II	Yes	Yes	No	Yes	Prospective	No missing data	Mortality	External	2801
Meyfoidt 2011[47]	Meyfroidt	Yes	Yes	No	No	Not specified	Imputed	LOS-ICU	Internal	499
Heldwein 2011[37]	LODS	Yes	Yes	No	Yes	Prospective	No missing data	Mortality	External	2801
Badreldin 2012[43]	SOFA, CASUS	Yes	Yes	No	Yes	Prospective	No missing data	Mortality	External	2801
	SOFA	Yes	Yes	No	Yes	Prospective	No missing data	Mortality	External	2801
Doerr 2012[44]	CASUS	Yes	Yes	Yes	Yes	Prospective	No missing data	Mortality	External	4054
Doerr 2014[29]	SAPS-II, SAPS -III	Yes	Yes	No	Yes	Prospective	No missing data	Mortality	External	5207
Ariyaratnam 2015[3]	APACHE-II, ICNARC	Yes	Yes	Yes	Yes	Prospective	Not discussed	Mortality	External	1646
Exarchopoulos 2015[24]	APACHE-II, SAPS-II, SOFA, CASUS	Yes	Yes	No	Yes	Prospective	No missing data	Mortality	External	150
Tsaousi 2015[25]	APACHE-II, SOFA	Yes	Yes	No	Yes	Prospective	No missing data	Mortality	External	1058

APACHE-II – Acute Physiology and Chronic Health Evaluation-II, APACHE-III – Acute Physiology and Chronic Health Evaluation-III, SAPS-II – Simplified Acute Physiology Score II, MODS – Multiple Organ Dysfunction Score , SOFA – (Sepsis-Related) Sequential Organ Failure Assessment, LODS – Logistic Organ Dysfunction Score, ICURS – Intensive Care Unit Risk Stratification Score, SAPS-3 – Simplified Acute Physiology Score 3, CASUS – Cardiac Surgery Score, ICNARC – Intensive Care National Audit and Research Centre, ICU Intensive Care Unit, AKICS – Acute Kidney Injury after Cardiac Surgery

LOS-ICU – Length Of Stay on the Intensive Care Unit, AKI – Acute Kidney Injury

Figure 1

Figure 1 -Manuscript selection for review