PARADISE: Predicting Atrial Fibrillation after Cardiac Surgery - the PARADISE Score. A Clinical Prediction Rule for Postoperative Atrial Fibrillation in Patients Undergoing Cardiac Surgery

Statistical Analysis Plan

Based on version 5 (3 July 2023) of protocol

ClinicalTrials.gov: NCT05255224

Ethics Ref: 21/EE/0166

CAG Ref: 21/CAG/0097

IRAS Project ID: 296508

Joint Research Office Ref: PID15669

HTA Project: NIHR131227

Lead statistician

Name: Professor Gary Collins

Signature:

Date:

01-March-2024

TABLE OF CONTENTS

1. FORWARD

This document details the proposed statistical analysis and presentation of the results for the main paper reporting the findings from the *NIHR funded Clinical Prediction Rule for Postoperative Atrial Fibrillation in Patients Undergoing Cardiac Surgery after Cardiac Surgery -the PARADISE Score.* This plan is intended to establish the rules that will be followed, as closely as possible, when modelling and reporting the prediction model.

This analysis plan will be available on request after submitting the main papers for publication in a scientific journal. Any deviations from the statistical analysis plan will be described and justified in the final report of the study. An identified, appropriately qualified, and experienced statistician will conduct the analysis, as well as ensuring the integrity of the data during their processing. Examples of such procedures include quality control and evaluation procedures.

1.1. KEY PERSONNEL

Statistician(s):
Jie Ma
Centre for Statistics in Medicine, Nuffield Department of Orthopaedics,
Rheumatology and Musculoskeletal sciences
jie.ma@csm.ox.ac.uk

Professor Gary Collins
Centre for Statistics in Medicine, Nuffield Department of Orthopaedics,
Rheumatology and Musculoskeletal sciences
Phone: 01865 223463
gary.collins@csm.ox.ac.uk

Chief Investigator:
Professor Peter Watkinson
Kadoorie Centre, University of Oxford, Level 3, John Radcliffe Hospital,
Headley Way, Oxford, OX3 9DU
Phone: 01865 857613
peter.watkinson@ndcn.ox.ac.uk

Lead Investigator (Joint Lead Applicant):
Professor Benjamin O'Brien
Perioperative Medicine Barts Heart Centre
Ben.OBrien@dhzc-charite.de

Investigator(s):
Kara Gayle Fields
Harvard Medical School
kgfields@bwh.harvard.edukg

Dr Oliver C Redfern
Kadoorie Centre, University of Oxford, Level 3, John Radcliffe Hospital,
Headley Way, Oxford, OX3 9DU
oliver.redfern@ndcn.ox.ac.uk

2. BACKGROUND

Atrial Fibrillation after cardiac surgery (AFACS) is the most common complication following cardiac surgery, with an incidence between 30% and 50%. The risk is higher for patients having either isolated mitral valve or combined mitral valve and coronary artery bypass graft surgery [1, 2]. Approximately 35,000 patients undergoing cardiac surgery in the UK every year. In the United States, approximately 500,000 patients undergo cardiac surgery annually [3]. AFACS is strongly associated with adverse patient outcomes, longer hospital and ICU stays, increased risk of stroke, increased risk of developing long-term AF (with associated complications and need for anticoagulation) and increased all-cause 30-day and 6-month mortality [4-9].

Even though AFACS can be transient, and patients are often discharged from hospital in normal sinus rhythm, patients with new-onset AFACS have a 5-fold increased risk of developing long-term AF [1]. A number of risk factors have been identified, including (older age, obesity) [10, 11], comorbidities (e.g., prior AF, hypertension, chronic kidney disease, obstructive sleep apnoea) [12], surgical features (e.g., valve surgery, increased aortic cross-clamp time, acute kidney injury) [13, 14] and the failure or inability to (re-)introduce Beta-Blockers. However, each of these factors only accounts for part of the overall risk [15]. Importantly, even when controlling for these confounding variables are, long-duration AFACS (> 2 days) is independently associated with decreased survival [16], as well as stroke and respiratory complications [17]. Current evidence therefore indicates that AFACS itself contributes to poor patient outcomes following cardiac surgery, and that tools to predict, prevent and guide treatment of AFACS are needed.

Currently, there is no widely accepted prediction model that reliably allows clinicians to determine the risk of a patient developing AFACS, despite multiple efforts over the past 15 years to develop one [18-21]. Published AFACS prediction models are limited by small sample sizes, a failure to include modern variables and a lack of external prospective validation [22-25]. Moreover, our team and others have previously developed AFACS risk prediction models that suffer from limited generalizability, oversimplification, and the fact that they often use information from variables that occur after surgery, or even after onset of AFACS [15, 24, 26].

This study will develop and validate two different scores as they apply to different situations, in the preoperative assessment clinic (PARADISE 1) and on arrival in the post-operative care unit (PARADISE 2), have different input variables (the latter including intra-operative and immediately available postoperative variables) and different time horizons for prediction (from weeks before the operation and from immediately post-operative in the post-operative care unit). These two models will have two distinct use cases. For example, PARADISE-1 could be used to start prophylactic medication in the preoperative assessment clinic, whereas PARADISE 2 will be used to adjust risk on admission to the postoperative care unit. Of the patients who develop AFACS, 70% do so before post-operative day 4 and 94% by post-operative day 7 [6, 27]. Therefore, both models will be optimised to predict AFACS within the first week of the initial post-operative hospital stay. This ensures we address the unmet need for prediction of AF following cardiac surgery, rather than AF in the context of prolonged hospitalisation or critical illness.

2.1. AIM

This study aims to develop and validate two clinical prediction models to estimate the risk of a patient developing AF in the seven days following cardiac surgery, using data available:

- In the pre-operative assessment clinic or at admission for surgery (PARADISE-1)
- On arrival to the post-operative care unit (PARADISE-2)

2.2. OBJECTIVES

To develop and internally validate two prognostic models to predict postoperative atrial fibrillation after cardiac surgery using data available in the pre-operative assessment clinic or at admission for surgery (PARADISE-1); and data available on arrival in the postoperative care unit (PARADISE-2) from a single US centre (Massachusetts General Hospital).

To externally validate the developed models on three separate datasets: a US hospital (Brigham and Women's Hospital), a UK clinical trial (Tight K study) and prospectively collected data from UK NHS cardiac surgery centres.

3. METHODOLOGY

This is an international, multi-centre longitudinal cohort study of patients who have undergone cardiac surgery. The study will use retrospective data to develop clinical prediction models to estimate the risk of developing atrial fibrillation in the seven days following surgery and use prospectively collected data to externally validate the developed models.

3.1. STUDY POPULATION

3.1.1. INCLUSION CRITERIA

- Patients 18 years or over
- Who underwent or are scheduled to undergo cardiac surgery

3.1.2. EXCLUSION CRITERIA

- Patients who have requested that their data not be used for research
- Patients with paced rhythm
- The procedure was solely:
    - A transcatheter insertion of a valve or stent
    - An Isolated vascular procedure (aorta, pulmonary artery)
    - A heart transplant
    - A lung transplant
    - For congenital heart disease

3.2. OUTCOME MEASURE

A new onset atrial fibrillation was defined as present if current atrial fibrillation or flutter was mentioned in patient cardiac report (e.g., ECG, ECHO) OR atrial fibrillation was identified by a trained chart abstractor for The Society of Thoracic Surgeons Adult Cardiac Surgery Database during postoperative period during the hospital admission (the outcome was assumed to happen within 7 days when timing was not specified).

Prior AF status was defined as any previous history of AF. Models for prior AF patients will be investigated in the sub analysis (section 3.6)

3.3. MODEL DEVELOPMENT

The following procedures will be applied for the development of the two prognostic models. Information from all subjects in the Massachusetts General Hospital database will be used to develop the two predictive models proposed in this analysis plan. Both models will predict the same outcome (Atrial Fibrillation) after cardiac surgery but will differ in the predictors they use to predict the outcome.

3.3.1. STUDY SETTING AND DATA COLLECTION

We will use the Massachusetts General Hospital (MGH) database to develop the PARADISE prediction models. The MGH dataset has two data sources:

1. the Research Patient Data Registry (RPDR): medical record data extracted from electronic hospital systems
2. the Society of Thoracic Surgeons (STS) Adult Cardiac Surgery Database: data gathered by trained extractors

Both data sources for the Mass General Hospital dataset will be used to develop our models and we will use retrospective pre-, intra-, and post-operative data for over 10,000 cardiac surgeries performed and recorded at this single US centre between 2003 and 2020.

It is important to note that some tables from this dataset were affected by a change in its electronic medical record system, which changed from Longitudinal Medical Record (LMR) to EPIC in April 2016.

3.3.2. Sample size calculation

As the two models to be developed are predicting the same outcome (new onset AF in the 7 days after surgery), the same sample size is applicable to both models.

Data on approximately 12,000 surgeries are available from Massachusetts General Hospital between 2003 – 2020 with approximately 20% of patients having new onset AF in the 7 days after surgery.

We used formal sample size formulae for binary outcome measurement provided by Riley et al [28] and the associated 'pmsampsize'[29] sample size calculator in R statistical software programme to calculate the maximum number of candidate predictor parameters we can include during model development. Based on the prevalence and assuming an anticipated R squared between 5% and 15% of the max Cox-Snell R2 (i.e. Cox-Snell R2 between 0.0316 and 0.0948) we can examine up to 42 - 132 predictor parameters for inclusion in the model [28].

3.3.3. Predictor selection before modelling

Prior to model development, we used the following diverse range of approaches to identify a comprehensive list of candidate predictors for AFACS:

- **Literature – systematic literature review:** We will perform a systematic review and critical appraisal of pre-operative variables predictive of AFACS. We will include existing prediction models and associated publications, both for AFACS and other AF substrates, such as primary AF (not after intervention), AF in critically ill patients and post-operative AF after non-cardiac surgery.
- **Data-driven identification – CALIBER database:** We will use patient data within the CALIBER research platform to identify risk factors (comorbidities, medical conditions, and medications) prior to cardiac surgery associated with incident AF.
- **Expert panel – modified Delphi process:** We will select a panel of international experts from relevant publications identified in our systematic review. The panel will undertake a modified Delphi process to identify a list of potential risk factors for AFACS.
- **Machine learning methods – exploring combinations of risk factors in the CALIBER and PRD Databases:** We will combine all candidate variables identified from the systematic review, CALIBER analysis and Delphi into a combined "long-list". Any variables missing from the PRD (to be used for model development) will be extracted from the electronic patient record. We

will then apply machine learning methods to the CALIBER and PRD data sets to derive combinations of risk factors (i.e., Interactions) that should be considered for inclusion in model development.

- **Expert panel – consolidating risk factors to be used in model development:** At their final meeting, the expert panel will review the final longlist of candidate variables (and their combinations). Any variables judged to be redundant or unlikely to represent spurious correlations will be removed.

The final candidate predictors are described in the study Case Report Form v2.1.

3.3.4. During modelling

In the Massachusetts General Hospital dataset, we anticipate that around 18% of patients will experience AFACS which was used to inform the sample size calculation. We will model the binary outcome using logistic regression. Though the risk of overfitting is low, given the high number of candidate predictors, we will also use shrinkage methods and specifically the Least Angle Selection and Shrinkage Operator (LASSO) to penalise and identify which predictors will be included in the final models [30].

Continuous variables will be kept as continuous in the model to avoid a loss of predictive information. Non-linear relationships with AFACS will be investigated using fractional polynomials or restricted cubic splines.

Since the development cohort spans an extended period, there is a chance that the calibration of the model will vary according to time. Therefore, we may consider re-calibrating the final model using the most recent subset of data [31].

3.3.5. Assumption checks

The normality of residuals will be checked using graphical methods. Influence of individual data points will be assessed by plotting leverage residuals against fitted data.

3.3.6. Machine Learning approaches

In parallel to develop statistical prediction models, we will investigate machine learning (ML) approaches to developing a prediction model, including methods such as random forests, deep neural networks, and Bayesian Gaussian processes. Such models allow consideration of interactions between available variables providing “latent variables” (complex, non-linear transformations of the original input variables) that may improve prediction performance over the non-transformed input variables. We will prioritise principled, probabilistic methods that permit the incorporation of prior clinical knowledge, such that results are “interpretable”, avoiding the “black box” nature of much ML-oriented research in this area [32, 33].

We will exploit the ability of the complementary modern ML fields of (i) Bayesian non-parametric modelling (which provides a principled means of coping with artefact and measurement noise in the clinical data) and (ii) deep learning, which permits the fusion of large quantities of (potentially time-varying) clinical variables. These models include components (“attention mechanisms”) that allow the resulting risk estimations to be interpretable for clinicians.

Importantly, these models permit the model to quantify its uncertainty in its score prediction – where conventional medical statistical methods always output a score, ML models provide the opportunity to quantity probabilistically the certainty with which a score is produced; this offers advantages in applications in which classes are substantially overlapped, as often occurs in complex, real-world clinical problems, where there is imperfect class separation. Accuracy of such models can often be substantially improved by permitting a model not to make a classification if the score is highly uncertain; such approaches are used in many application areas of critical systems to improve acceptability with users.

Additionally, we will investigate the joint construction of phenotypical clustering models (which identify reproducible phenotypes in the physiological data) with predictive models specific to each phenotypical group – this novel approach constructs an ensemble of predictive models that are phenotype-specific, thereby improving predictive accuracy with respect to a single "one-size-fits-all" generic algorithm applied to the entire population.

3.4. General issues for statistical analysis

3.4.1. Method for handling missing data

To avoid excluding patients when developing and validating our models, we will use multiple imputation to impute missing values, under a missing at random assumption.

Identifying the true underlying missing data mechanism from the available is rarely possible. Assumptions need to be made on the plausible mechanism, and approaches needed to be used. Under a missing completely at random mechanism. (MCAR), the missingness mechanism does not depend on unobserved (unseen) data. Conducting a complete case analysis will produce unbiased estimates (but with a loss precision if full data are observed). Under the missing at random (MAR) assumption, the missingness after conditioning on the observed data does not depend on the unobserved (unseen) data. Under this approach, we can apply approaches such as multiple imputation, by fitting a joint model to the observed data and impute the missing data, taking account of the uncertainty in the estimated parameters of this joint model.

We feel this, MAR, approach makes a less strong and more realistic assumption than the MCAR approach. The MAR imputation model will include all variables considered for the multivariable model building, the outcome and any auxiliary variables that will help explain the missingness. The assumption of a missing not at random (MNAR) approach whilst not implausible is considerably more complex to investigate – there is a dearth of research investigating MNAR in the context of prediction model research. We will nevertheless explore whether the MAR assumption holds by comparing the imputed values (after accounting for the observed values) and the missing values to identify if there are any systematic differences to suggest a MNAR assumption.

Therefore, to conform to current guidelines[34], multiple imputation for all subjects with at least one missing value (using the 'mice' function in R, for example) will be used.

Data from LHCH and BHC provide a large "real-world" prospective validation but is likely to contain more missingness [35].

Using sensitivity analyses we may investigate the impact of informative missingness in our model development, i.e. departures from the missing at random assumption.

3.4.2. METHOD FOR HANDLING OUTLIERS

Outliers will be identified by plotting box plots of each continuous variable. Clinical judgment will be used to assess if the outlier could be a true value.

3.5. INTERNAL VALIDATION

After developing a prognostic model, it is important to evaluate its performance. Internal validation can be performed using methods applied to the same data from which the model was developed. Internal validation of the final model will also be assessed by the bootstrap resampling technique using at least 200 iterations whereby each modelling step is repeated in each bootstrap sample, to adjust for over-optimism in the estimation of model performance. ML models will also be additionally validated via k-fold cross-validation. The internal validation will quantify and be used to adjust the performance measures (e.g., discrimination, calibration) for any optimism.

3.5.1. MODEL PERFORMANCE

The performance of a prediction model is typically assessed by its calibration and discrimination.

Calibration, which reflects how close the predictions from the model are to the observed outcome frequencies will be assessed graphically, using a calibration plot, plotting observed outcomes against predictions using smoothing techniques. The calibration plot will be supplemented with a smoothed regression line. Calibration will also be quantified by calculating the calibration slope and intercept.

The discrimination of the prediction models is the ability of the model to correctly rank individuals (i.e., those with the event should have higher predicted probabilities than those without), will be summarised with the concordance index (equivalent to the Area Under Receiver Operating Characteristic curve) with 95% confidence interval.

3.6. SUB ANALYSIS

An additional analysis will be investigated to include patients with prior AF. Models (both regression model and machine learning approaches) will be developed to population include prior AF patient (compared to population for patients exclude prior AF). Model coefficients and performance will be compared between patients with and without prior AF.

3.7. CLINICAL UTILITY/NET BENEFIT

Decision curve analysis will be used to assess and compare the utility of the models (both regression model and machine learning approaches) and to explore at 25% of threshold which calculates the net benefit of using the model (compared to not using the model). The net benefit of a model is the difference between the proportion of true positives and the proportion of false positives weighted by the odds of the selected threshold for high-risk designation [36].

3.8. EXTERNAL VALIDATION

3.8.1. Study setting and data collection (validation datasets)

The developed models will be validated on three datasets. These are:

Brigham and Women's Hospital: The Brigham and Women's CABG Genomics Database is a prospective single centre and detailed research database comprising over 2000 surgeries from 2001-2016. This dataset contains over 1,700 prospectively collected and curated pre, intra, and post-operative variables.

NHS Trusts: The study will prospectively collect data from the UK NHS Trusts which will allow us to undertake external validation of PARADISE-1/PARADISE-2 with a total of approximately 6000 patients (2000 AFACS events) per annum. The NHS Trusts where data will be extracted from are:

- Liverpool Heart & Chest NHS Foundation Trust
- Barts Heart Centre (Barts Health NHS Trust)
- Oxford University Hospitals NHS Foundation Trust

Tight K study (trial dataset): (BHF-funded CS/18/3/34063, ClinicalTrials.gov Identifier: NCT04053816). Tight K is a 22-centre clinical trial that will include 1684 patients, randomised within this non-inferiority trial to investigate whether maintaining serum potassium ≥3.6mEq/L is equivalent to maintaining levels at 4.5- 5.5mEq/L, with a primary endpoint of development of AFACS using precise clinical definitions and 5-day Holter monitoring. The trial is already designed to collect detailed data on AFACS risk factors, minimising any additional burden of PARADISE. It is restricted to patients undergoing isolated coronary artery bypass grafting. This has the strength that detailed data are being collected in a well-defined cohort, but means we need additional prospective data to assess the performance of the PARADISE scores in valve surgery.

3.8.2. Sample size calculation

Current recommendations for external validation sample size calculation uses 3 criteria (outlined below) [37]. We based the observed events/expected events (O/E) on the outcome prevalence (20%), the calibration slope (assuming model is well calibrated, intercept=0, slope=1), and on the c-statistic (15% of the max Nagelkerke's R-squared).[37-39]. We have assumed the linear predictor is normally distributed with a common variance in event and non-event groups, prior to model development. Given this strong assumption, we will update these calculations after model development is completed.

1. O/E calculation
   We assume O/E is 1 in the external validation population, with a 95% confidence interval width of 0.2 for O/E. he estimated sample size to satisfy criterion 1 is 1538 patients (~308 events).
2. Calibration slope
   We assume model is well calibrated with intercept=0 and slope=1. We also assume the linear predictor distribution is normally distributed. For a 95% confidence interval width of 0.2, we would need a sample size of 4248 patients (~850 events).
3. C-statistic
   Based on 15% of R^2 (i.e., max 15% Cox-Snell R2 of ~0.09 based on assumed prevalence),

which equates to a C-statistic of ~0.72. Using this estimated C-statistic value with targeted a SE(C) of 0.025, correspond to a confidence interval of width 0.1 (i.e., 0.67-0.77), we would need a sample size of 619 patients (about 124 events). Varying the expected C-statistic from 0.67 to 0.77 and based on the assumption above, a sample size of 531 patients (106 events) to 691 patients (138 events) would be needed, respectively.

Therefore, a minimum sample size of 4248 patients with 850 (20% of 4248) events would be needed for external validation to satisfy all three criteria.

Both regression models and machine learning approaches will then be externally validated on prospective data from one US hospital (Brigham and Women's Hospital), prospectively collected data from UK NHS Trusts and one UK clinical trial (Tight K study). In brief, for each individual in these datasets, outcome predictions will be calculated using the model developed with the Massachusetts General Hospital dataset and compared with the observed outcomes. The performance of the prediction models will be characterised by evaluating calibration and discrimination. To account for potential differences in case-mix (distribution of patient characteristics and prevalence of the outcome) between the US data (Massachusetts General Hospital dataset) used to develop the model and the prospective external validation, we will investigate whether recalibration is needed to update the model to better fit the UK population [35].

Heterogeneity in model performance will be explored over different hospitals using internal-external cross-validation [40, 41].

Finally, we will compare our prediction models to those existing prediction models identified from the systematic review where sufficient details are available from the original publication and predictors are available in the US data set.

4. REPORTING

The description of the development and internal validation of the prediction models will be reported according to the TRIPOD and statement [34] and TRIPOD+AI should be available at the time for ML approaches[42].

5. REFERENCES

1. Lowres, N., et al., *Incidence of postoperative atrial fibrillation recurrence in patients discharged in sinus rhythm after cardiac surgery: A systematic review and meta-analysis.* Interactive cardiovascular and thoracic surgery, 2017. **26**.
2. Greenberg, J.W., et al., *Postoperative atrial fibrillation following cardiac surgery: a persistent complication.* European Journal of Cardio-Thoracic Surgery, 2017. **52**(4): p. 665-672.
3. Badhwar, V., et al., *The Society of Thoracic Surgeons Adult Cardiac Surgery Database: 2018 Update on Research: Outcomes Analysis, Quality Improvement, and Patient Safety.* Ann Thorac Surg, 2018. **106**(1): p. 8-13.
4. Almassi, G.H., et al., *Postoperative atrial fibrillation impacts on costs and one-year clinical outcomes: the Veterans Affairs Randomized On/Off Bypass Trial.* Ann Thorac Surg, 2015. **99**(1): p. 109-14.
5. Aranki, S.F., et al., *Predictors of atrial fibrillation after coronary artery surgery. Current trends and impact on hospital resources.* Circulation, 1996. **94**(3): p. 390-7.
6. Gialdini, G., et al., *Perioperative atrial fibrillation and the long-term risk of ischemic stroke.* Jama, 2014. **312**(6): p. 616-22.
7. Horwich, P., K.J. Buth, and J.F. Légaré, *New onset postoperative atrial fibrillation is associated with a long-term risk for stroke and death following cardiac surgery.* J Card Surg, 2013. **28**(1): p. 8-13.
8. LaPar, D.J., et al., *Postoperative Atrial Fibrillation Significantly Increases Mortality, Hospital Readmission, and Hospital Costs.* The Annals of Thoracic Surgery, 2014. **98**(2): p. 527-533.
9. Shen, J., et al., *The persistent problem of new-onset postoperative atrial fibrillation: a single-institution experience over two decades.* J Thorac Cardiovasc Surg, 2011. **141**(2): p. 559-70.

10. Phan, K., et al., *Obesity and postoperative atrial fibrillation in patients undergoing cardiac surgery: Systematic review and meta-analysis.* Int J Cardiol, 2016. **217**: p. 49-57.
11. Wong, C.X., et al., *Obesity and the Risk of Incident, Post-Operative, and Post-Ablation Atrial Fibrillation: A Meta-Analysis of 626,603 Individuals in 51 Studies.* JACC Clin Electrophysiol, 2015. **1**(3): p. 139-152.
12. Qaddoura, A., et al., *Obstructive sleep apnea as a predictor of atrial fibrillation after coronary artery bypass grafting: a systematic review and meta-analysis.* Can J Cardiol, 2014. **30**(12): p. 1516-22.
13. Cole, O.M., et al., *Acute Kidney Injury and Postoperative Atrial Fibrillation In Patients Undergoing Cardiac Surgery.* J Cardiothorac Vasc Anesth, 2020. **34**(7): p. 1783-1790.
14. Indja, B., et al., *New Onset Atrial Fibrillation Following Transcatheter and Surgical Aortic Valve Replacement: A Systematic Review and Meta-Analysis.* Heart Lung Circ, 2020. **29**(10): p. 1542-1553.
15. Pollock, B.D., et al., *Predicting New-Onset Post-Coronary Artery Bypass Graft Atrial Fibrillation With Existing Risk Scores.* Ann Thorac Surg, 2018. **105**(1): p. 115-121.
16. Sigurdsson, M.I., et al., *Duration of Postoperative Atrial Fibrillation After Cardiac Surgery Is Associated With Worsened Long-Term Survival.* Ann Thorac Surg, 2016. **102**(6): p. 2018-2026.
17. Megens, M.R., L. Churilov, and V. Thijs, *New-Onset Atrial Fibrillation After Coronary Artery Bypass Graft and Long-Term Risk of Stroke: A Meta-Analysis.* J Am Heart Assoc, 2017. **6**(12).
18. Amar, D., et al., *Clinical prediction rule for atrial fibrillation after coronary artery bypass grafting.* J Am Coll Cardiol, 2004. **44**(6): p. 1248-53.
19. Mathew, J.P., et al., *A multicenter risk index for atrial fibrillation after cardiac surgery.* Jama, 2004. **291**(14): p. 1720-9.

20. Thorén, E., et al., *Prediction of postoperative atrial fibrillation in a large coronary artery bypass grafting cohort.* Interact Cardiovasc Thorac Surg, 2012. **14**(5): p. 588-93.
21. Zaman, A.G., et al., *Atrial fibrillation after coronary artery bypass surgery: a model for preoperative risk stratification.* Circulation, 2000. **101**(12): p. 1403-8.
22. El-Chami, M.F., et al., *Prediction of new onset atrial fibrillation after cardiac revascularization surgery.* Am J Cardiol, 2012. **110**(5): p. 649-54.
23. Helgadottir, S., et al., *Atrial fibrillation following cardiac surgery: risk analysis and long-term survival.* J Cardiothorac Surg, 2012. **7**: p. 87.
24. Kolek, M.J., et al., *Genetic and clinical risk prediction model for postoperative atrial fibrillation.* Circ Arrhythm Electrophysiol, 2015. **8**(1): p. 25-31.
25. Tran, D.T., et al., *Predicting New-Onset Postoperative Atrial Fibrillation in Cardiac Surgery Patients.* J Cardiothorac Vasc Anesth, 2015. **29**(5): p. 1117-26.
26. Cameron, M.J., et al., *Prospective External Validation of Three Preoperative Risk Scores for Prediction of New Onset Atrial Fibrillation After Cardiac Surgery.* Anesth Analg, 2018. **126**(1): p. 33-38.
27. Echahidi, N., et al., *Mechanisms, prevention, and treatment of atrial fibrillation after cardiac surgery.* J Am Coll Cardiol, 2008. **51**(8): p. 793-801.
28. Riley, R.D., et al., *Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes.* Statistics in Medicine, 2019. **38**(7): p. 1276-1296.
29. Joie Ensor, E.C.M., Richard D. Riley, *Calculates the Minimum Sample Size Required for Developing a*

Multivariable Prediction Model. 2022.
30. Tibshirani, R., *Regression and shrinkage via the Lasso.* J R Stat Soc, Ser B, 1996. **58**(Journal Article): p. 267-288.

31. Booth, S., et al., *Temporal recalibration for improving prognostic model development and risk predictions in settings where survival is improving over time.* Int J Epidemiol, 2020. **49**(4): p. 1316-1325.
32. Shamout, F.E., et al., *Deep Interpretable Early Warning System for the Detection of Clinical Deterioration.* IEEE J Biomed Health Inform, 2020. **24**(2): p. 437-446.
33. Yang, Y., et al., *DeepAMR for predicting co-occurrent resistance of Mycobacterium tuberculosis.* Bioinformatics, 2019. **35**(18): p. 3240-3249.
34. Collins, G.S., et al., *Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis: The TRIPOD statement.* Ann Intern Med, 2015. **162**: p. 55-63.
35. Fanaroff, A.C., et al., *Stroke prevention in atrial fibrillation: re-defining 'real-world data' within the broader data universe.* Eur Heart J, 2018. **39**(32): p. 2932-2941.
36. Vickers, A.J. and E.B. Elkin, *Decision curve analysis: a novel method for evaluating prediction models.* Med Decis Making, 2006. **26**(6): p. 565-74.
37. Riley, R.D., et al., *Minimum sample size for external validation of a clinical prediction model with a binary outcome.* Statistics in Medicine, 2021. **40**(19): p. 4230-4251.
38. Snell, K.I.E., et al., *External validation of clinical prediction models: simulation-based sample size calculations were more reliable than rules-of-thumb.* Journal of Clinical Epidemiology, 2021. **135**: p. 79-89.
39. Riley, R.D., B. Van Calster, and G.S. Collins, *A note on estimating the Cox-Snell R(2) from a reported C statistic (AUROC) to inform sample size calculations for developing a prediction model with a binary outcome.* Stat Med, 2021. **40**(4): p. 859-864.
40. Riley RD, E.J., Snell KIE, Debray TPA, Altman DG, Moons KGM, et al., *External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges.* BMJ, 2019. **365**: p. l4379.

41. Takada, T., et al., *Internal-external cross-validation helped to evaluate the generalizability of prediction models in large clustered datasets.* Journal of Clinical Epidemiology, 2021. **137**: p. 83-91.
42. Collins, G.S., et al., *Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence.* BMJ Open, 2021. **11**(7): p. e048008.