

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Thuraisingam, S;Dowsey, M;Manski-Nankervis, J-A;Spelman, T;Choong, P;Gunn, J;Chondros, P

Title:

Developing prediction models for total knee replacement surgery in patients with osteoarthritis: Statistical analysis plan

Date:

2020-12

Citation:

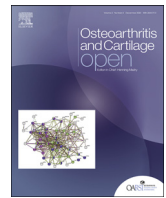
Thuraisingam, S., Dowsey, M., Manski-Nankervis, J. -A., Spelman, T., Choong, P., Gunn, J. & Chondros, P. (2020). Developing prediction models for total knee replacement surgery in patients with osteoarthritis: Statistical analysis plan. *Osteoarthritis and Cartilage Open*, 2 (4), pp.100126-100126. <https://doi.org/10.1016/j.ocarto.2020.100126>.

Persistent Link:

<https://hdl.handle.net/11343/258820>

License:

[CC BY-NC-ND](#)



Developing prediction models for total knee replacement surgery in patients with osteoarthritis: Statistical analysis plan

Sharmala Thuraisingam^{a,*}, Michelle Dowsey^a, Jo-Anne Manski-Nankervis^b, Tim Spelman^{a,c}, Peter Choong^a, Jane Gunn^d, Patty Chondros^b

^a Department of Surgery, University of Melbourne, 29 Regent Street, Fitzroy, Victoria 3065, Australia

^b Department of General Practice, University of Melbourne, 780 Elizabeth Street, Parkville, Victoria 3010, Australia

^c Karolinska Institute, Solnavagen 1, 171 77 Solna, Sweden

^d Faculty of Medicine Dentistry & Health Sciences, Level 2, Alan Gilbert Building, Carlton, Victoria 3053, Australia

ARTICLE INFO

Keywords:

Prediction models
Clinical prediction tools
Statistical analysis plan
Electronic medical record
Electronic health record
Knee replacement
General practice
Primary care

SUMMARY

Background: Approximately 12–20% of those with osteoarthritis (OA) in Australia who undergo total knee replacement (TKR) surgery do not report any clinical improvement. There is a need to develop prediction tools for use in general practice that allow early identification of patients likely to undergo TKR and those unlikely to benefit from the surgery. First-line treatment strategies can then be implemented and optimised to delay or prevent the need for TKR. The identification of potential non-responders to TKR may provide the opportunity for new treatment strategies to be developed and help ensure surgery is reserved for those most likely to benefit. This statistical analysis plan (SAP) details the statistical methodology used to develop such prediction tools.

Objective: To describe in detail the statistical methods used to develop and validate prediction models for TKR surgery in Australian patients with OA for use in general practice.

Methods: This SAP contains a brief justification for the need for prediction models for TKR surgery in general practice. A description of the data sources that will be linked and used to develop the models, and estimated sample sizes is provided. The planned methodologies for candidate predictor selection, model development, measuring model performance and internal model validation are described in detail. Intended table layouts for presentation of model results are provided.

Conclusion: Consistent with best practice guidelines, the statistical methodologies outlined in this SAP have been pre-specified prior to data pre-processing and model development.

1. Background

Osteoarthritis (OA) affects over two million Australians (9% of the population) and over 30% of those aged 65 years and over [1]. In 2016, 181 knee and 112 hip replacements per 100,000 population were performed, contributing to a total estimated healthcare cost of \$2.1 billion for the management of OA [2,3]. This is forecasted to reach \$2.9 billion by 2030 [2,4–6]. Despite this, approximately 12–20% of those who undergo TKR surgery do not report improvements in pain and function [7,8].

General practitioners (GPs) are often the first health professional that patients consult for treatment advice, usually for joint pain or lack of function. They play a critical role in assisting patients to find the most appropriate management options. Whilst guidelines recommend non-surgical and non-pharmacological interventions as first-line for patients with OA, uptake is reported as low [9]. The reasons for this are complex and may be due to characteristics of the patient, intervention and/or health system [10]. Early identification of patients likely to undergo TKR in the future provides a great window of opportunity for first-line

Abbreviations: ABS, Australian Bureau of Statistics; AIHW, Australian Institute of Health and Welfare; AOANJRR, Australian Orthopaedic Association National Joint Replacement Registry; ATC, Anatomical Therapeutic Chemical; BMI, Body Mass Index; CPT, clinical prediction tool; DQA, data quality assessment; EMR, electronic medical record; GP, General Practitioner; KOS-ADLS, Knee Outcome Survey-Activities of Daily Living Subscale; NDI, National Death Index; NPS, National Prescribing Service; OA, osteoarthritis; OMERACT, Outcome Measures in Rheumatology; OARSI, Osteoarthritis Research Society International; SAP, statistical analysis plan; SF-36, 36-Item Short Form Health Survey; SF-12, 12-Item Short Form Survey; TKR, total knee replacement.

* Corresponding author.

E-mail addresses: sharmala.thuraisingam@unimelb.edu.au (S. Thuraisingam), mmdowsey@unimelb.edu.au (M. Dowsey), jomn@unimelb.edu.au (J.-A. Manski-Nankervis), tim@burnet.edu.au (T. Spelman), pchoong@unimelb.edu.au (P. Choong), j.gunn@unimelb.edu.au (J. Gunn), p.chondros@unimelb.edu.au (P. Chondros).

<https://doi.org/10.1016/j.ocarto.2020.100126>

Received 16 July 2020; Accepted 17 November 2020

2665-9131/© 2020 Osteoarthritis Research Society International (OARSI). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

treatment strategies to be implemented and optimised to suit the patient, with the possibility of preventing or delaying the need for TKR in the future. Identification of patients least likely to respond to TKR provides the opportunity for new treatment strategies to be developed and trialled for this small but important subset of patients. Redirection of these patients to new therapies may help ensure resources for TKR surgery are reserved for those most likely to benefit.

To date, there are no prediction models specifically developed for use early in the patient's osteoarthritis journey in a clinical setting such as primary care. We conducted a literature review to Ref. [1]: identify and critically appraise prediction models developed for TKR and response/non-response to TKR and [2] identify factors predictive of or associated with TKR and response/non-response to TKR. A summary of the findings is detailed in [Appendix A](#).

Of the 30 studies identified in the literature review, four related to predicting TKR [11–14], six related to predicting response/non-response to TKR [8,15–19] and the remainder reported associations with TKR (12 studies) [20–31] and response/non-response to TKR (eight studies) [32–39].

Of the studies predicting TKR, only one reported on the performance of the model (c-statistic = 0.79) [11], only one addressed missing data [14], and the statistical methods used to develop the models were not well reported in two of the studies [11,12]. Factors found to be predictive of TKR included measures of knee pain and physical function, use of medication for knee pain, Knee Outcome Survey-Activities of Daily Living Subscale (KOS-ADLS), 36-Item Short Form Health Survey (SF-36) general health subscale score, willingness to undergo TKR, seeing a healthcare provider for arthritis, knee osteoarthritis grade and 12-Item Short Form Survey (SF-12) mental component score [11–14]. Other factors found to be associated with TKR not identified in the prediction models included weight gain between middle age and early adulthood, change in joint space width, duration and intensity of leisure time physical activity, BMI, change in BMI, gout, interaction between gout and BMI, interaction between BMI and age, smoking status and physical activity at work [20–31].

For the second outcome, response/non-response to TKR, model performance was reported in two studies (c-statistics = 0.74 [8] and c-statistic = 0.77 [18]). The statistical methods used for model development

Table 1
Variables to be used in model development.

Variable	Data source	Variable type	Coding
Model one			
Time to TKR (outcome)	AOANJRR	Continuous	Days
Status	AOANJRR	Binary	0 = Did not undergo TKR 1 = Underwent TKR 2 = Died
Age	NPS EMRs	Continuous	Years
BMI	NPS EMRs	Continuous	Kg/m ²
Prescribing of medication/s for OA	NPS EMRs	Binary	1 = Prescribed medication for osteoarthritis 0 = Not prescribed medication for osteoarthritis
Multimorbidity	NPS EMRs	Continuous (discrete)	Count of chronic conditions listed in: - Charlson Comorbidity Index (CCI) [28] - BEACH study [30] - Combination of CCI and BEACH [28,30]
Mental health condition	NPS EMRs	Binary	1 = Diagnosis of anxiety, depression, post-traumatic stress disorder (PTSD), obsessive compulsive disorder (OCD), anorexia, bulimia, bi-polar disorder, dissociative disorders or schizophrenia 0 = None of the conditions listed above
Weight gain between early adulthood and middle age	NPS EMRs	Continuous	Difference in weight at 45–65 years and 18–21 years
Previous/contralateral knee replacement	AOANJRR	Binary	1 = Underwent contralateral/previous TKR 0 = Did not undergo contralateral/previous TKR
Any past knee surgery on either knee	NPS EMRs	Binary	1 = Underwent arthroscopy, open reduction/repair, reconstruction, cruciate ligament repair, clean out, debridement, fracture of tibial plateau with screws, supracondylar fracture with femur pin, periprosthetic fracture of femoral condyle, meniscus repair, meniscectomy, lateral release, anterior cruciate ligament repair, medial collateral ligament repair, osteotomy, chondroplasty, fracture tibial plateau with repair, avulsion fracture of femoral condyle or arthrotomy prior to study 0 = Did not undergo any of the surgeries listed above
Geographical location (surgeon access)	NPS EMRs	Categorical	As per Australian Statistical Geography Standard (ASGS) remoteness areas: 0 = Major cities 1 = Inner regional 2 = Outer regional 3 = Remote 4 = Very remote
Model two			
Non-response to TKR (outcome)	SMART	Binary	1 = Non-responder to TKR (as per OMERACT OARSI criteria) 0 = Responder to TKR
BMI	SMART	Continuous	Kg/m ²
Multimorbidity	SMART	Continuous (discrete)	Count of chronic conditions listed in: - Charlson Comorbidity Index (CCI) [28] - BEACH study [30] (17 most commonly managed chronic conditions in primary care reliably recorded in EMR) - Combination of CCI and BEACH [28,30]
Mental health condition	SMART	Binary	1 = Diagnosis of depression, anxiety, bipolar disorder, schizophrenia, post-traumatic stress disorder, borderline personality disorder or obsessive-compulsive disorder 0 = None of the conditions listed above
Use of opioids prior to surgery	MBS/PBS	Categorical	0 = Use of non-opioid osteoarthritis medication only 1 = Use of opioid osteoarthritis medication only 2 = Use of both non-opioid and opioid osteoarthritis medication
Fibromyalgia	SMART	Binary	1 = Diagnosis of fibromyalgia 0 = No diagnosis of fibromyalgia

were generally well described in five out of the six studies [8,15–18], however, only one study addressed missing data [16]. Age, severity of osteoarthritis, number of other troublesome hips and knees, body mass index (BMI), comorbidities, SF-12 mental component score, measures of pain and function, and presence of low back pain were found to be predictive of response to TKR surgery [8,15–19]. Other factors associated with response to TKR not identified in the prediction models included fibromyalgia, pain relief expectations and psychological distress [32–39].

All models identified in the literature review for both TKR and response to TKR were based on predictors collected in hospital, randomised controlled trials or allied health clinics and therefore unlikely to be applicable in the general practice setting. Hence, there is a need for two clinical risk prediction tools (CPTs) to be developed from data routinely collected in general practice to identify early those patients likely to undergo TKR in the future, and those patients least likely to respond to TKR.

The focus of this research and statistical analysis plan (SAP) is to develop two CPTs for TKR in patients with OA aged 45 years and over. The first will predict the likelihood of TKR in the next five years and the second model will predict the probability of non-response to TKR one-year post-surgery. These CPTs will be translated into web-based applications that are embedded within the general practice EMR so that GPs have access to these tools at the point of care.

2. Objectives

To develop prediction models for use in Australian general practice for patients aged 45 years and over with OA to predict:

- (1) the likelihood of TKR in the next five years (model one)
- (2) the probability of non-response to TKR one-year post-surgery (model two)

3. Outcomes to be predicted

For model one, the outcome measure is time to primary TKR. For model two, the outcome measure is non-response to primary TKR one-year post-surgery. Non-response to TKR surgery will be determined using the Outcome Measures in Rheumatology-Osteoarthritis Research Society International (OMERACT OARSI) responder criteria [8,40]. The coding of these variables is outlined in Table 1.

4. Methods

4.1. Predictor selection

We used the approach recommended in Steyerberg et al. [41] for predictor selection. This involved a literature review and consultation with clinical experts. An adapted Delphi process [42–44] was then used to obtain consensus amongst experts on potential predictive factors (results in Appendix B). Given the models will be used in general practice and embedded in the EMR, the list of potential predictors was further reduced to include only those routinely collected in general practice and hence available in the EMR (Table 1). Predictors identified in the literature and by clinical experts that were excluded are listed in Appendix B.

4.2. Data sources and patient population

4.2.1. Model one

The primary data source for the development of model one will be patient electronic medical records (EMRs) from Australian general practices. This data set has been provided by MedicineWise as part of the MedicineInsight data program and consists of approximately 475,870 consenting patients from 671 practices across Australia with a recorded diagnosis of OA in their EMR [45]. Data include all patient data entered in the EMR by the December 31, 2017 (inclusive) except for administrative data (e.g. dates of consultation) which has only been provided for 2013–2017.

The study cohort will be limited to “active” patients, defined as patients attending the clinic at least twice during 2013. This is to ensure the EMR data used in model development is up to date and representative of patient characteristics at study baseline (2014). Fig. 1 outlines the timelines for model one. Given the outcome is time to first TKR, patients who underwent bilateral TKR prior to baseline will be excluded.

All predictors and the outcome measures will be derived from the MedicineInsight general practice EMRs. However, due to uncertainty in the accuracy of recording of events in EMRs that occur outside of the primary care setting such as TKR surgery and death, these data fields will be validated against gold standard national registries, the Australian Orthopaedic Association National Joint Replacement Registry (AOANJRR) [46] and National Death Index (NDI) [47] respectively via data linkage. Data relating to TKR date and date of death from these national registries will be considered accurate and used in model

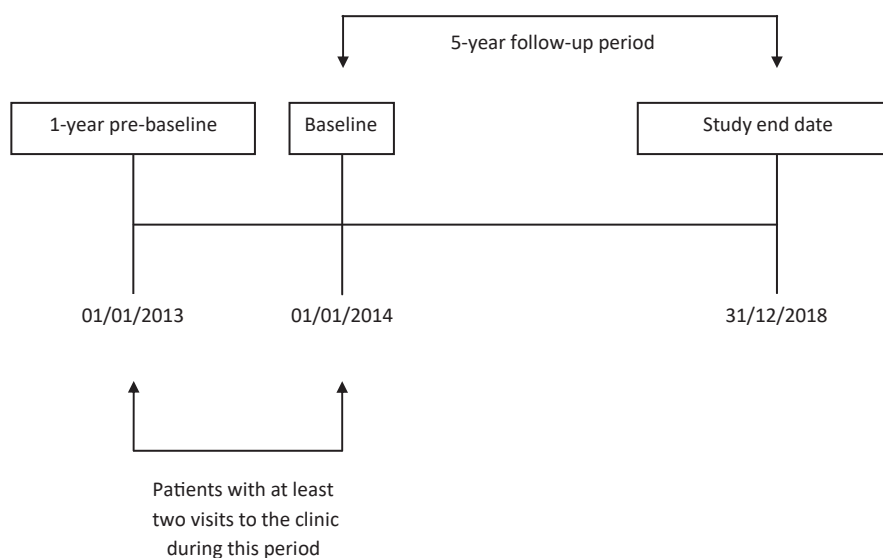


Fig. 1. Study timeline for model one.

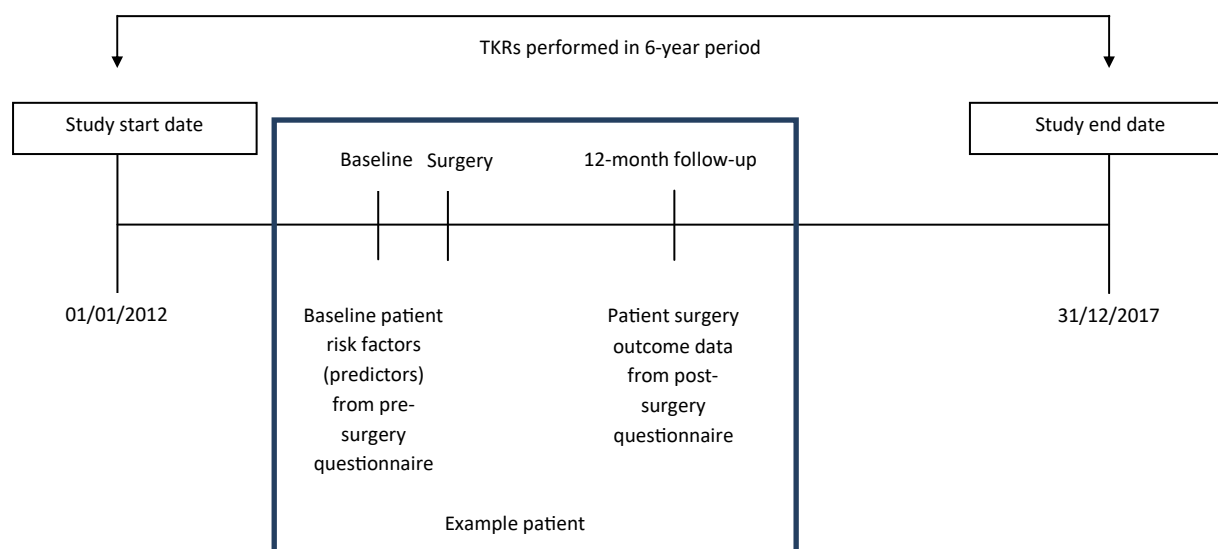


Fig. 2. Study timeline for model two.

development should these data be inaccurate or incomplete in the EMR.

4.2.2. Model two

The predictors and outcome for model two will be obtained from the St Vincent's Melbourne Arthroplasty Outcomes Registry (SMART) [8]. The SMART dataset contains pre- and post-surgery data from approximately 6800 patients who underwent TKR at St Vincent's Hospital in Melbourne since 1998. Data are most complete post 2012. Therefore, the study period for model two will be the January 1, 2012 to the December 31, 2017 (Fig. 2). Here, baseline will be defined as the date of collection of patient pre-surgery characteristics in SMART which is typically performed as close as possible to surgery. Follow-up is 12 months post-surgery.

Linkage with Pharmaceutical Benefits Scheme (PBS) data from the Australian Institute of Health and Welfare (AIHW) will be used to obtain data on pain medication use which is not routinely recorded in SMART.

4.3. Data linkage

Figs. 3 and 4 depict the planned data linkage methodology for each of the models. The data linkage methodology is described in detail in Appendix C.

4.4. Sample size

There is no standard way of calculating sample sizes for prediction models [41,48]. However, with approximately 15,000 TKRs for model one, the events per variable based on a model with nine predictors will be over 1500. For model two, no clinical benefit was reported for approximately 19% of the 2400 TKRs. With five predictors for this model, the events per variable is approximately 90. This is sufficient for the development of stable models [41].

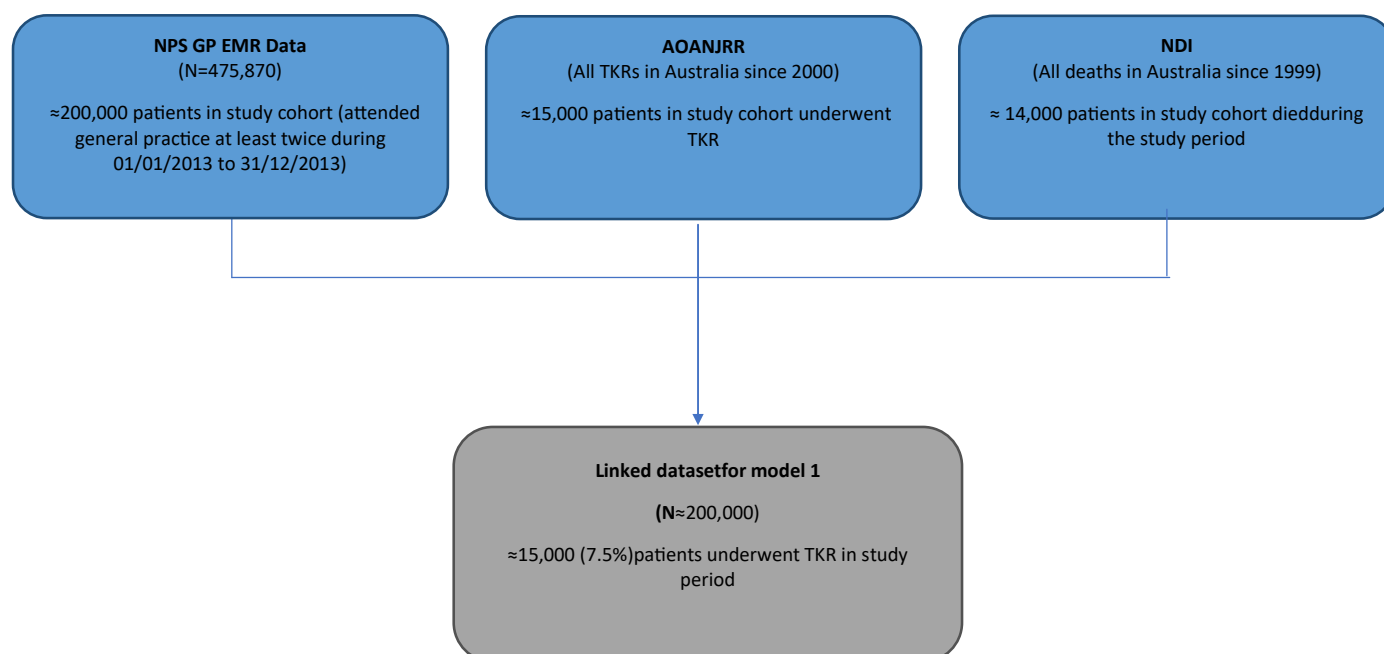


Fig. 3. Data linkage for model one.

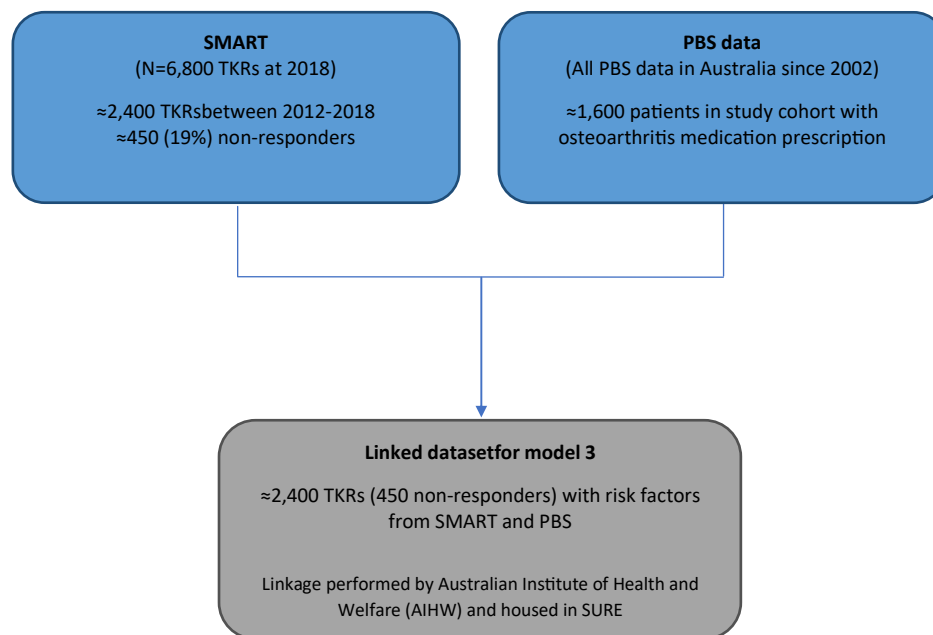


Fig. 4. Data linkage for model two.

4.5. Data pre-processing

Whilst EMRs are a rich source of data, they were designed for clinical and administrative purposes and not specifically for research. As a result, these records are rarely in the format required for analyses and extensive data re-configuration and pre-processing is often required.

Table 1 summarises characteristics of the variables that will be used in model development. Since there is no widely accepted global measure for overall health that is routinely collected in general practice, multimorbidity count will be used as a proxy measure. We will consider three different ways of counting multimorbidity: (i) count of chronic conditions listed in the Charlson Comorbidity Index (CCI) [49], (ii) count of 17 frequently managed chronic conditions in primary care as identified in the BEACH study [50], and (iii) a combination of (i) and (ii). Further details of the coding and pre-processing of these variables is provided in Appendix D.

4.6. Data quality assessment

Data cleaning and pre-processing does not guarantee that the data will be fit for the intended purpose. Therefore, prior to building our prediction models, we will assess the quality of the available data to determine its suitability for reliably answering the research question. The data quality assessment (DQA) of the EMR and SMART data sets will follow the methods proposed by Kahn et al. [51] and will be published in a separate paper. The results of this undertaking will inform whether the EMR and SMART data sets are suitable for our intended purpose. Should we find the data quality acceptable, we will proceed with developing the two models.

4.7. Model development

The following sections outline the prediction modelling methodology that will be used to develop the models. The seven step modelling approach proposed by Steyerberg et al. [41] will be followed (Table 2). STATA version 15.1 (StataCorp) [52] will be used for model development.

4.7.1. Data inspection

Descriptive statistics will be used to summarise the characteristics of patients. Missing data will be assessed by summarising the amount of missing data for each predictor, identifying whether non-missing predictor values are correlated with the missingness of other predictors and auxiliary variables, and identifying associations between missingness and the outcome [41]. Provided the proportion of missing data for predictor variables does not exceed 35% and missing data can be explained by the data available, multiple imputation (MI) using chained equations will be performed to impute missing values for the predictors [53–56]. The approach is outlined below. Complete cases will be used instead [54], when the missing data of all the predictors included in the model is less than 10%.

4.7.1.1. Model one. Imputation models will include all candidate predictors, the outcome, and the frequency of attendance at general practice as this is known to be associated with missing data [41,55]. Any other reasons for missing data that are identified from considering the context in which the data were collected, collated and distributed, and that can be explained by variables within the data will be included as auxiliary variables in the imputation models [41,55].

The number of imputations will exceed the percentage of missing data [54]. For non-normal continuous predictors, imputations will be performed on the transformed scales [54]. The convergence of the imputation process will be assessed by plotting the mean and standard deviation of the imputed values across iterations. The fit of the imputation models will be assessed by comparing the distributions of the observed and imputed data. If imputation model convergence is achieved and model fit is adequate, prediction models will be developed using the imputed data sets.

There should be no missing outcome data for model one given that recordings of TKR in the EMR will be validated against data from the AOANJRR.

4.7.1.2. Model two. Given the rigorous annual data checks carried out on the SMART data set to correct missing and implausible data in the registry, missing data are expected to be less than 10% for model two

Table 2
Prediction model development checklist.

	Model one	Model two
Data inspection	Descriptive statistics, missing data patterns and multiple imputation (if missing at random)	
Coding of predictors	Levels of categorical variables will be combined if cell counts are less than 10 and/or the model becomes unstable Non-linear transformations considered for continuous variables Truncation considered for possible influential outliers	
Model specification	Fine and Gray proportional hazards Selection of main effects: Full model with removal of predictors with HR/OR between 0.90 and 1.10 and p-value>0.1	Logistic regression
Model estimation	Maximum log-likelihood and shrinkage (bootstrapping)	
Model performance	Discrimination- Harrell's overall C statistic Calibration- predicted and observed 5-year risk	Discrimination- Concordance C statistic Calibration- predicted and observed probabilities of non-response
Model validation	100 bootstrap samples of each imputed data set to assess model optimism External validation outside scope of this project	
Model presentation	Regression equation, nomogram and web-based tool (future)	

predictors. Hence MI will not be performed for model two.

Given that accurate predictions and the relationship between predictors and the outcome is of primary interest, and not mean prevalence of the outcome, methods for dealing with missing outcome data will not be adopted here and only TKRs with complete follow-up data will be included in model two [41]. However, a comparison of baseline characteristics between patients with missing and complete follow-up data will be carried out to explore potential indicators of study cohort selection bias.

4.7.2. Coding of predictors

The coding of predictors has been summarised previously and is further detailed in Appendix D. Levels of categorical variables will be collapsed with adjacent categories if cell counts are less than 10 and/or the model becomes unstable when the predictor is fitted to the model [57–59]. For model two, the cell count will be obtained through cross-tabulation with the outcome. Non-linear transformations for continuous predictors and truncation for potentially influential observations will be considered.

4.7.3. Model specification

A Fine and Gray proportional hazards competing risk regression model [60,61] will be used for model one, with death treated as a competing risk. Continuous predictors will be centred. To account for the clustering of patients within clinics, robust variance estimates will be calculated [62,63]. The proportional hazards assumption will be checked for each categorical variable using the first imputed data set by inspecting plots of the survival distribution function against the survival time for parallel lines, and by testing for interactions between each predictor and time [64]. An interaction between age and BMI was found to be associated with TKR in Leyland et al. [25] and therefore will be tested during model one development.

For model two, logistic regression will be used [65] with robust variance estimation to account for correlation in the outcome for patients contributing data for two TKRs (left and right).

We intend to use MI using approximately 30 imputed data sets and validate our models internally using 100 bootstrap samples drawn randomly with replacement from each imputed data set [66] ($30 \times 100 = 3000$ data sets). Initially, a full model will be fit to each bootstrapped sample within each imputed data set. Predictors with subdistribution hazards (model one) or odds (model two) ratios between 0.90 and 1.10 and p-values>0.1 will be excluded from the model [41,62,67]. The inclusion frequency will be defined for each predictor as the proportion of times the predictor is selected in each of the bootstrapped samples. Models based on threshold frequencies of 60%, 70%, 80% and 90% will then be developed in each of the imputed data sets. Rubin's rules [68] will be used to average regression estimates across imputed datasets for each of the threshold frequencies and presented as shown in Table 3.

Table 3

Model regression estimates by threshold frequency.

Threshold	Subdistribution Hazard ratio (95% CI)	P-value
Model one		
60%		
70%		
80%		
90%		
Model two		
60%		
70%		
80%		
90%		

Abbreviations: CI = confidence interval.

4.7.4. Model estimation

Regression coefficients will be estimated through maximisation of the log-likelihood function. A correction factor for optimism will be determined from bootstrapping and applied to the regression coefficients.

4.7.5. Model performance

Harrell's overall C statistic for model one and the concordance statistic C for model two will be used to assess model discrimination [41]. To assess calibration, the slopes of the predicted and observed 5-year likelihood of TKR (model one) and probability of non-response to TKR (model two) will be considered.

4.7.6. Model validation

The bootstrapped samples will be used to obtain optimism corrected estimates of model discrimination and calibration. The optimism corrected calibration slopes will then be used to obtain shrinkage factors for regression estimates [41,66]. This process is detailed in Appendix E.

The final model will be selected by comparing the optimism corrected model performance and calibration slopes for each threshold frequency [66]. Results will be presented as shown in Tables 4 and 5.

External model validation will be carried out as a separate study and is not reported in this SAP.

4.7.7. Model presentation

The final models will be presented as regression equations and converted into nomograms [69]. Each predictor value in the nomogram will correspond to a regression weight such that totals are equivalent to the linear predictor. For model one, the predicted 5-year survival estimate will be obtained by exponentiating the linear predictor and multiplying by the baseline subdistribution hazard function at 5 years. For model two, the logistic transformation will be applied to the linear predictor to produce estimates of the probability of non-response to TKR.

Table 4
Model performance measures by threshold frequency.

Threshold	Number of predictors	Apparent C-statistic	Optimism corrected C-statistic (from bootstrapping)	Mean slope of linear predictors (from bootstrapping)
Model one				
60%				
70%				
80%				
90%				
Model two				
60%				
70%				
80%				
90%				

Table 5
Regression coefficients for final models with shrinkage (from bootstrapping).

Model one	Subdistribution Hazard ratio (95% CI)	P-value
Model two	Odds ratio (95% CI)	P-value

Abbreviations: CI = confidence interval.

5. Discussion

This SAP outlines the statistical methodology that will be used to develop the two prediction models. As with any planned analysis, there are limitations. For model one, we do not have access to medication dispensing data, only prescriptions issued by general practitioners. Therefore, we cannot be certain that specific osteoarthritis medication issued by the general practitioner was dispensed and taken by patients. Whilst this may seem problematic at first, when we consider that the models will be used in the general practice setting, it seems appropriate to develop the models purely from data (predictors) recorded in the EMR. Further, the medications considered in this study are not only used for OA and it is possible that patients are taking these medications for other purposes.

For model two, preliminary data linkage between the EMR data and patient follow-up data from SMART did not yield significant numbers of records for model development. Hence, predictors will need to be derived from baseline data within SMART. Whilst we have only selected candidate predictors that are available in the general practice EMR we understand that differences in these data sources, such as the reasons for data collection and data collection methods may result in slightly different models had the data been developed using EMRs. Differences may be greatest for the predictor relating to opioid medication, where dispensed OA medication from the PBS will be used as a proxy for prescribed OA medication due to limitations in the data available.

Lastly, our choice of predictor selection methodology was somewhat restricted due to the lack of established methods in the context of multiply imputed data sets with bootstrap validation. Further research into the use of more modern predictor selection techniques such as LASSO in this context is needed.

There are several strengths in our prediction modelling methodology. Firstly, the development is driven by the context in which the prediction tools will be used, and the methodology has been developed with input from clinical experts to ensure our models are clinically useful, relevant and pragmatic. For example, potential end-users of the prediction tools prefer models that will clearly quantify the effects of each predictor on the outcome and include only clinically meaningful predictors. This is consistent with the requirements of practitioners documented in the

literature [70]. For this reason, regression modelling techniques were selected over machine learning.

Another strength of this methodology is the use of data linkage with gold-standard registries (AOANJRR and NDI) to obtain accurate outcome data for model development. The quality of outcome and predictor variables, and external validity of our study cohorts will be assessed in a structured data quality assessment prior to model development.

The methods documented in this SAP are well established and address some of the limitations within the data. For example, factors incorporated within the MI models will address the context in which these data were collected and some of the possible biases that may arise along the way. Further, bootstrapping will be used to provide stable estimates of model optimism and to study the stability of predictors, and our models will be developed from data sets with a large number of outcome cases.

The use of primary care EMRs for research and developing prediction models is in its infancy in Australia. This important work will inform how these data may be used to develop clinically useful prediction tools.

Declarations

Ethics approval and consent to participate

We have obtained the necessary ethics approvals from the following: The University of Melbourne ID 1852593, St Vincent's Hospital ID 46036 LRR 202/18, The University of South Australia ID 201840 and Australian Institute of Health and Welfare EO2018/5/509.

The analyses are based on secondary data sources and consent to use the data for this research purpose has been granted by participants (SMART, AOANJRR) or general practice clinics (MedicineInsight EMR data). This study has been approved by the NPS MedicineWise data governance committee.

Consent for publication

Not applicable.

Availability of data and material

All data governance procedures have been followed for each of the respective data sets and approvals to use the data for the intended research purpose has been provided by each of the data providers.

Competing interests

The authors declare that they have no competing interests in relation to this study.

Funding

Funding for part of this study has been provided by the RACGP Foundation/HCF Research Foundation Grant from the Royal Australian College of General Practitioners (RACGP). Funding has also been provided by the Centre for Research Excellence in Total Joint Replacement.

Author's contributions

ST drafted and finalised the statistical analysis plan with critical input from PChondros, MD, JMN and TS. PChondros and TS assisted with the development of the analysis methodology. All authors contributed to the final manuscript.

Acknowledgements

We would like to acknowledge NPS Medicinewise for providing the primary care EMR data, the Australian Orthopaedic Association for the joint registry data and the Australian Institute of Health and Welfare for the National Death Index. We would also like to acknowledge BioGrid Australia for facilitating and performing the data linkage for this project and the Australian Institute of Health and Welfare for linking the EMR data with the NDI.

This work is supported by the National Health and Medical Research

Council of Australia (NHMRC) Centre for Research Excellence in Total Joint Replacement (APP1116325). Sharmala Thuraisingam is the recipient of a scholarship awarded through the NHMRC Centre for Research Excellence in Total Joint Replacement (APP1116235). Michelle Dowsey holds a NHMRC Career Development Fellowship (1122526) and a university of Melbourne Dame Kate Campbell Fellowship. Peter Choong holds a NHMRC Practitioner Fellowship (1154203). Jo-Anne Manski-Nankervis holds a Medical Research Future Fund Next Generation Clinical Researchers Program - Translating Research into Practice Fellowship (1168265).

Appendix A. Summary of literature review

Background and Aim:

We conducted a literature review to Ref. [1]: identify and critically appraise prediction models developed for TKR and response/non-response to TKR and [2] identify factors predictive of or associated with TKR and response/non-response to TKR. The following section includes the methods used in the literature search, the overall findings of the studies identified, a critical summary of the methodologies used to develop the prediction models and concludes with identification of factors to consider during model development.

Methods:

Medline, EMBASE and CINAHL databases were searched using the criteria documented in Table A 1.1. These databases were chosen due to their relevance to the search topic and are consistent with recommendations by the US Agency for Healthcare Research and Quality (AHRQ) and UK Centre for Reviews and Dissemination (CRD) for more rigorous systematic reviews [71]. Gray literature such as conference abstracts and theses were also included in the review to reduce the risk of publication bias. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram [72] summarising the search results are displayed in Figure A 1.1. Given the type of literature under review, the Critical Appraisal Skills Programme (CASP) Clinical Prediction Rule Checklist [73] was used as a framework to critically appraise the studies. Whilst meta-analysis can be used to summarise the effect of predictors on a particular outcome across studies, it may be subject to publication bias as many studies only report results for statistically significant predictors [41]. Meta-analysis was therefore not included as part of this review.

Table A 1.1
Literature review search criteria

(AND) →			
(OR)↓	MeSH (knee replacement/arthroplasty) knee replacement* knee arthroplast*	MeSH (osteoarthritis/knee osteoarthritis) knee osteoarthritis	MeSH (models/statistical) model* MeSH (nomograms) nomogram*

Limit: English language and last 10 years.

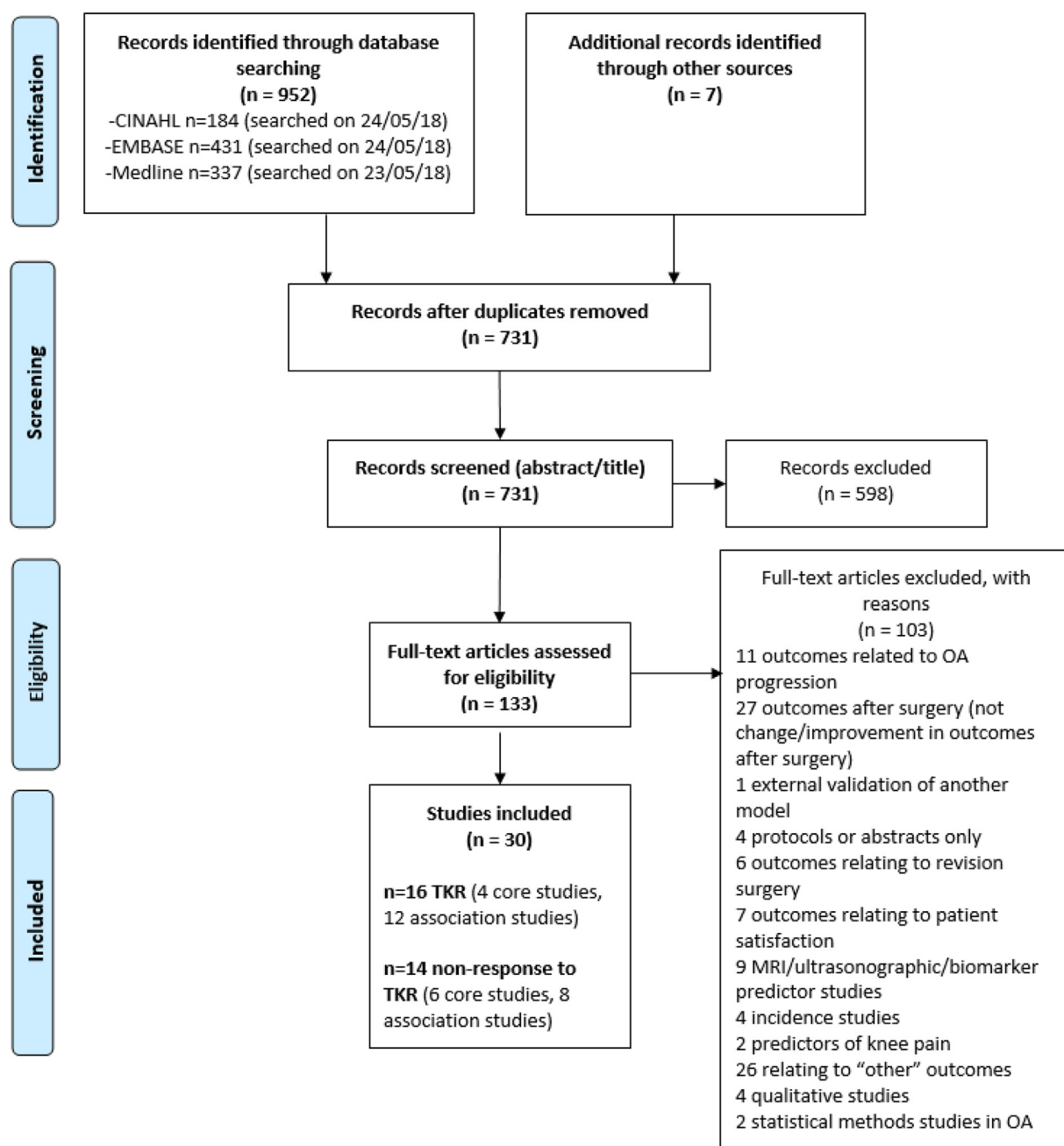


Fig. A 1.1. PRISMA flow diagram

Results:

Prediction models and studies identified

From the literature search, 30 studies in total were identified [8,11–31]. Four included prediction models for TKR [11–14] and six included prediction models for response/non-response to TKR [8,15–19]. These are the core papers that were appraised in this review. Of the remaining 20 studies, 12 examined associations between pre-specified factors of interest and TKR [20–31], and eight examined associations with response/non-response to TKR [32–39]. As these studies were not considered to be core papers, they were not critically appraised. However, they provided insight into other possible predictive factors that may not have been included in the core papers and were therefore used to answer aim two of the literature review.

Aim 1: Critical appraisal of prediction models

Table A 1.2 summarises the findings from applying the Critical Appraisal Skills Programme (CASP) Clinical Prediction Rule Checklist [73] to the prediction models identified for TKR and response/non-response to TKR.

Table A 1.2
Summary of critical appraisal

Criteria	Model 1: Predictors of TKR	Model 2: Predictors of response/non-response to TKR
Is the prediction model clearly defined?	All four studies clearly defined the outcome, risk prediction interval, predictors included and cohort the model would apply to.	All six studies clearly defined the outcome, risk prediction interval, predictors included and cohort the model would apply to. However, Jiang et al. [15] and Clement et al. [18] defined any change in OKS score 12 months post-surgery to be indicative of positive response to surgery. Small changes in OKS score may not be clinically significant. Lungu et al. [17] defined poor responders to TKR at 6 months as those with WOMAC scores in the last quintile. This may not necessarily equate to poor response.
Has the method for predictor selection been clearly defined?	Overall, the methods used for selecting prediction factors for potential inclusion in the models were poorly documented. In the five studies that did provide some justification, the drivers for factor selection included previous factors identified in the literature, factors easily measurable in the respective clinical setting and factors identified by experts with clinical knowledge [14,17,36–38]. In the papers that did not provide insight into factor selection methodology it is possible that this was influenced by the factors available within their data set as majority of the studies utilised pre-existing data as opposed to conducting a study.	The populations were well chosen in four out of the six studies. Clement et al. [18] did not provide any information on the study population except that they were patients identified as having undergone TKR in 2007–2009. Lungu et al. [17] selected patients on the waiting list for TKR in 3 hospitals in Quebec City, however, only 141 patients were selected.
Did the population from which the rule was derived include an appropriate spectrum of patients?	Acceptable in three out of the four studies. Zeni et al. [36] used a homogenous sample of patients with radiographic evidence of OA who were referred to a physical therapy clinic by one orthopaedic surgeon (who does not perform TKR) in Delaware, USA. This model is unlikely to be applicable outside of this specific setting.	
Was the rule validated in a different group of patients similar to the one used to derive it?	None of the prediction models were validated using external data sets.	
Were the predictor variables and outcome evaluated in a blinded fashion?	All predictors of TKR were measured at baseline and the occurrence of TKR either tracked throughout the duration of the study [14, 36] or ascertained by linking data from hospital records [11,37]. WOMAC and OKS scores for model 2 were obtained at either 6 months, 12 months or 5 years post-surgery. The predictor variables and outcomes were therefore likely to be measured and ascertained independent of each other, reducing the likelihood of model development from bias data. However, the exact details of how this occurred was not clearly documented in any of the studies.	
Are exclusions and drop-outs well described and reasonably treated?	Not all studies reported exclusions and drop-outs well. Exclusions were well described in Lewis et al. [11], Riddle et al. [14], Dowsey et al. [38] and Lungu et al. [17]. Hawker et al. [13] and Jiang et al. [15] reported drop-outs and compared the characteristics of the drop out cohort to the study cohort. Clement et al. [18] did not detail drop-outs or exclusions. Razak et al. [16] reported drop-out but did not do any further analysis or testing.	
Are the statistical methods used to construct and validate the rule clearly defined and methodologically sound?	The statistical methods used by Hawker et al. [37] and Riddle et al. [14] were generally well described. However, Hawker et al. [37] did not detail the modelling steps from attainment of the univariable results to the development of the final model. Riddle et al. [14] was the only study to address missing data. The statistical methods used by Lewis et al. [11] and Zeni et al. [36] were not well reported.	The statistical methods were generally well described in five out of the six models. However, only Jiang et al. [15] addressed missing data by incorporating multiple imputation. The exact criteria used for stepwise model building required further explanation in most studies that used this approach. The statistical methods for model development used by Clement et al. [18] were not well described.
Can the performance of the rule be calculated/were performance measures provided?	Performance measures were only provided in two out of the four studies (Lewis et al. [11] and Zeni et al. [36]). Moderate model performance, C-statistic = 0.79 was reported in Lewis et al. [11]. Zeni et al. [36] did not report a C-statistic but did report sensitivity at 62% and specificity at 86% indicating that the model was better able to accurately predict those who do not undergo TKR.	Performance measures were only provided in two out of the six studies (Dowsey et al. [38] and Lungu et al. [17]). Moderate model performance was reported in both studies with a C-statistic of 0.74 in Dowsey et al. [38] and C-statistic of 0.77 in Lungu et al. [17].
How precise were the predictor effect estimates?	Precision of predictor effect estimates were acceptable in three out of the four studies. Zeni et al. [36] did not report confidence intervals and only 40 patients in this study had undergone TKR. Confidence intervals are expected to be wide for this study.	Precision of predictor effect estimates were acceptable in four studies (Dowsey et al. [9], Hawker et al. [13], Razak et al. [16] and Clement et al. [18]). The effect estimates of predictors was not provided in Jiang et al. [15] most likely due to non-significance. Lungu et al. [17] used a regression tree, and hence the precision of effect estimates is not applicable here.
Applicability of study results to our cohort	Both Lewis et al. [11] and Riddle et al. [14] include patients within the age group of our study (45 years and over) and hence results from these studies may be applicable. Hawker et al. [37] includes patients with TKR or THR and therefore results need to be interpreted with care. The results from Zeni et al. [36] are from one specific physical therapy clinic and orthopaedic surgeon in the United States and therefore unlikely to be applicable in our setting.	The results from Dowsey et al. [9], Hawker et al. [13], Jiang et al. [15] and Lungu et al. [17] are applicable to this research study due to similarities in the cohorts and follow-up times. The mean BMI of patients in the Razak et al. [16] study is likely to be lower than that of patients from Australia, however parallels can still be drawn with the remaining cohort characteristics, and results are therefore still applicable. Lastly, little information was provided on the cohort of patients used in Clement et al. [18] and it is difficult to determine the relevance of these results to this research study.

Aim 2: Identifying factors predictive of/associated with TKR and response/non-response to TKR

Factors predictive of/associated with TKR. There was little consistency in the factors considered and found to be predictive of TKR/total joint replacement (TJR) in the four studies (Table A 1.3). Differences may be due to the studies being conducted in different settings and one study modelling TJR as the outcome as opposed to TKR. Age and BMI were the only factors common in the four models. Increasing age was found to be predictive of TKR in Zeni et al. [12] and Riddle et al. [14], and of TJR in Hawker et al. (until 82 years) [13]. Only Lewis et al. [11] found BMI to be associated with TKR [11].

Table A 1.3

Summary of prediction models for TKR

Study	Lewis et al. (N = 1462)	Zeni et al. (N = 120)	Hawker et al. (N = 2128)	Riddle et al. (N = 4670)
Outcome	10 yr risk of TKR in Australian women ≥ 70 years with OA	TKR 2yrs post referral to orthopaedic surgeon in American patients ≥ 46 yrs with OA	5 yr risk of TJR in Canadian patients aged ≥ 55 years with OA of knee or hip	TKR at 3 yr follow-up in American patients aged 45–79 years with OA or at risk of OA
Setting	Data linkage study using hospital discharge data throughout Western Australia and data from a randomised controlled trial	University of Delaware Physical Therapy Clinic	Population study conducted in two regions in Ontario, Canada	Multicentre (1 hospital and 3 university sites) community-based study conducted in Maryland, Rhode Island, Ohio and Pennsylvania
Factors	Final model OR (95% CI)	Final model OR	Final model HR (95% CI)	Final model RR (95% CI)
Sociodemographic factors				
Age (years)	1.00 (0.93–1.07)	1.13**		1.04 (1.01–1.07)
≤62			1.00 (reference)	
63–68			1.57 (1.10–2.25)	
69–71			1.46 (1.01–2.10)	
75–81			1.51 (1.03–2.20)	
≥82			0.44 (0.22–0.88)	
Clinical factors				
BMI (kg/m ²)	1.07 (1.03–1.11)			
Knee osteoarthritis history				
Prevalent knee replacement/ past knee surgery	2.08 (1.01–4.28)			2.04 (1.33–3.13)
Previous hip replacement				2.73 (0.93–8.07)
Seeing a healthcare provider for arthritis				1.81 (1.12–2.92)
Knee OA grade (normal vs. severe)				2.09 (1.63–2.69)
Knee pain and symptoms				
Knee pain (infrequent, frequency, daily v.s no)	2.35 (1.86–2.97)			
Factors	Final model OR (95% CI)	Final model OR	Final model HR (95% CI)	Final model RR (95% CI)
WOMAC total score			1.22 (1.13–1.31)	
Medication for knee/joint pain	1.98 (1.33–2.95)			1.64 (0.87–3.12)
Physical function and examination				
eROM		1.23**		
Knee effusion bulge sign				1.58 (1.04–2.40)
Knee flexion contracture (degrees)				1.06 (1.02–1.11)
Lateral or medial tibiofemoral tenderness				0.71 (0.43–1.18)
Patient initiated knee flexion pain				1.58 (1.04–2.39)
Maximum knee extension/body weight (newtons/kg)				0.79 (0.65–0.96)
Repeated chair stand pace (stands/sec)				1.20 (1.04–1.40)
General health and daily living				
SF-36 general health subscale score			1.14 (1.07–1.21)	
SF-12 MCS score				1.07 (1.04–1.10)
KOS-ADLS		0.96**		
Considering all ways knee pain and arthritis affects you, how are you doing today? (0–10 scale)				1.12 (1.01–1.23)
Perceptions of TKR				
Willingness to undergo TKR			4.92 (3.73–6.44)	8.0 (4.32–14.83)

Abbreviations: Total Knee Replacement (TKR), Odds Ratio (OR), Confidence Interval (CI), Hazard Ratio (HR), Relative Risk (RR), Body Mass Index (BMI), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), extension Range of Motion (eROM), Short Form 36 (SF-36), Short Form 12 (SF-12), Mental Component Score (MCS), Knee Outcome Survey-Activities of Daily Living Subscale (KOS-ADLS).

** No confidence intervals provided but variable included in final model.

Measures of knee pain (pain frequency, pain severity and WOMAC pain score), were included in three out of the four prediction models. Frequency of knee pain and severity of knee pain were predictors of TKR in Lewis et al. [11] and Riddle et al. [14] respectively. Analgesia use for joint pain was only found to be predictive of TKR in Lewis et al. [11] despite also being included by Riddle et al. [14]. Hawker et al. [13] found the WOMAC total score to be a predictor of TJR suggesting that perhaps knee stiffness is predictive of TJR given that the pain and physical functioning subscales of this measure were not found to be predictors and WOMAC stiffness score was not individually tested. This conclusion is unclear due to lack of detail around the modelling process in this study, specifically the treatment of multicollinearity.

In Lewis et al. [11] the strongest predictors of TKR were knee pain and whether the patient had undergone contralateral knee replacement. Zeni et al. [12] found extension range of motion to best predict TKR of all the factors tested, whilst Hawker et al. [13] and Riddle et al. [14] found that patients

willing to consider joint replacement were almost five times and eight times more likely to undergo the procedure respectively. General health and mental health were predictive of TJR and TKR in Hawker et al. [13] and Riddle et al. [14] respectively.

Factors found to be associated with TKR that were not identified in the prediction models discussed previously included weight gain between early adulthood and middle age [20], change in joint space width (JSW) [21], gout [26], smoking [29] and physical activity at work [31]. An interaction between BMI and age was also identified with the odds of TKR greater for those aged less than 68 years compared with those aged 68 years or older [27].

Factors predictive of/associated with response/non-response to TKR. There were some similarities in the factors included as predictors in the six prediction models for non-response to TKR. These included age, gender and comorbidities (Table A 1.4). All studies except Clement et al. [19] included BMI, and all studies included some measure of mental health status and adjusted for baseline outcome. There were also several other factors considered solely by each of these studies. These included Kellgren-Lawrence (K-L) grade [8], number of other troublesome joints [15], presence of low back pain [15,19], socio-demographic [8,15–19] and surgery characteristics [15,17].

Table A 1.4

Summary of prediction models for response to TKR

Study	Dowsey et al. (N = 615)	Hawker et al. (N = 202)	Jiang et al. (N = 1636)	Razak et al. (N = 3062)	Lungu et al. (N = 141)	Clement et al. (N = 966)
Outcome	Non-response to TKR (defined using OARSI responder criteria) 12 months post-surgery in Australian patients	Response to TJR (defined as 0.5SD difference in WOMAC summary score) 12 months post-surgery in Canadian patients aged 55 years and older with at least moderately severe knee/hip arthritis	Response to TKR (defined as change in OKS score) 12 months post-surgery in UK patients	Good TKR outcome (defined as improvement in OKS of at least 0.5SD difference in preoperative and postoperative scores) 5 years post-surgery in Asian patients	Poor response to TKR (WOMAC score in last quintile) 6 months post-surgery in Canadian patients aged ≥ 40 years	Change in OKS 12 months post TKR in unspecified patient cohort with mean age 70.7 years
Setting	Data from an arthroplasty registry at a Melbourne hospital	Population study conducted in two regions in Ontario, Canada	Multicentre (34 centres) randomised controlled trial in the United Kingdom	Data from a tertiary institution in Singapore	Multicentre (3 hospitals) study in Quebec City, Canada	Unspecified
Factors	Final model OR (95%CI)	Final model RR (95% CI)	Final model Mean difference in change in OKS score (95% CI)	Final model OR (95% CI)	Final model (regression tree)	Final model Mean difference in change in OKS score (95% CI)
<i>Sociodemographic factors</i>						
Age				2.66 (2.61–2.71) [‡]		
<i>Arthritis history</i>						
KL grade (KL ≤ 3 v.s KL = 4)	2.59 (1.58–4.24)					
Etiology (inflammatory v.s OA)		0.33 (0.15–0.76)				
No. of other troublesome hips/knees		0.82 (0.72–0.93) [‡]				
<i>Clinical factors</i>						
Presence of low back pain						–2.53 (–3.75 to –1.30)
BMI (>40 kg/m ² v.s ≤ 40 kg/m ²)	3.48 (1.97–6.12)					
Comorbidities		0.88 (0.79–0.97) [‡]				–3.78 (–6.11 to –1.45) [‡]
<i>Measures of mental health</i>						
SF-12 MCS	1.00 no disability 0.93 (0.46–1.87) mild 1.74 (0.95–3.29) moderate 3.30 (1.44–7.58) severe					0.16 (0.11–0.22)
<i>Measures of function and pain</i>						
WOMAC summary score	0.81 (0.68–0.97)	1.32 (1.23–1.42) [‡]			*5 WOMAC questions [†]	
OKS			✓unspecified			*0.58 (0.50–0.67)
KSKS				2.64 (2.63–2.66) [‡]		
KSFS				2.66 (2.63–2.68) [‡]		

Abbreviations: Standard deviation (SD), Total Knee Replacement (TKR), Osteoarthritis Research Society International (OARSI), Osteoarthritis (OA), Kellgren and Lawrence (KL), Body Mass Index (BMI), Short Form-12 (SF-12), Mental Component Score (MCS), Short Form-36 (SF-36), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Oxford Knee Score (OKS), Knee Society Knee Score (KSKS), Knee Society Function Score (KSFS).

[‡]per 10 yr decrease in age.

[‡]per additional joint.

[‡]per additional condition up to 3.

[‡] ≥ 4 comorbidities v.s < 4 comorbidities.

[‡]per 10-point increase in score.

[†]No regression coefficients reported as tree model used.

Despite the similarities in factors selected across the studies, the findings were not consistent. Only Razak et al. [17] found decreasing age to be predictive of good TKR outcomes, and only Dowsey et al. [8] categorised BMI into morbidly obese/not morbidly obese ($\text{BMI} \geq 40 \text{ kg/m}^2$ compared with $< 40 \text{ kg/m}^2$) and found morbid obesity to be predictive of non-response to TKR. Other unique findings included decreased relative risk of good outcome post TKR with increasing number of troublesome joints (Hawker et al. [15]) and smaller changes in Oxford knee score (OKS) score post TKR in patients with low back pain (compared with no low back pain).

The differences in factors included in the prediction models may be due to differences in outcome definition across studies. For example, Dowsey et al. [8] modelled non-response to TKR 12 months post-surgery according to the Outcome Measures in Rheumatology-Osteoarthritis Research Society International (OMERACT) responder criteria. Hawker et al. [15] and Razak et al. [17] modelled response to TKR at 12 months and five years post-surgery respectively defining a positive response as at least 0.5 of a standard deviation change in WOMAC and OKS respectively. Jiang et al. [16] and Clement et al. [19] created models for change in OKS score 12 months post-surgery and Lungu et al. [18] modelled poor outcomes of TKR at 6 months post-surgery, defining poor responders as those with post-surgery WOMAC scores in the last quintile.

There were however, two consistent findings [1]; both Hawker et al. [15] and Clement et al. [19] found higher comorbidity count to be predictive of smaller changes in WOMAC and OKS scores respectively; and [2] Dowsey et al. [8] found greater odds of non-response to TKR in those with severe mental disability (compared to those without) and Clement et al. [19] found greater improvements in OKS with higher SF-12 mental component scores. Further, as expected, all studies found baseline outcome to be predictive of the final outcome measure.

Eight studies explored the association between patient factors and response/non-response to TKR. Opioid use up to two years before TKR [32], fibromyalgia survey score (measuring pain and comorbid symptoms) [33], pain relief expectations prior to surgery [34] and psychological distress as measured by the Hospital Anxiety and Depression Scale (HADS) [36] were found to be associated with response to TKR.

Summary of candidate predictors identified from the literature review

Table A 1.5 summarises the potential predictors for both models using all studies identified in the literature review. These predictors were proposed to a panel of experts in osteoarthritis (OA) using a Delphi process (results in Appendix B) to ensure all predictors to be used during model development are clinically relevant.

Table A 1.5
Candidate predictors identified from the literature

Model 1: Predictors of TKR	Model 2: Predictors of non-response to TKR
1. Age	1. Age
2. BMI	2. BMI
3. Seeing any healthcare provider for osteoarthritis	3. Number of other troublesome knees/hips
4. Knee osteoarthritis grade	4. Knee osteoarthritis grade
5. Frequency of knee pain	5. Number of comorbidities
6. Severity of knee pain	6. Mental health
7. Severity of knee stiffness	7. Presence of low back pain
8. Functional difficulty	8. Opioid use prior to surgery
9. Use of medication/s for osteoarthritis	9. Fibromyalgia
10. Extension range of motion	10. Pain relief expectations
11. Knee effusion bulge sign	
12. Knee flexion contracture (degrees)	
13. Patient initiated knee flexion pain	
14. Maximum knee extension force divided by body weight (newtons/kg)	
15. Repeated chair stand pace (stands/sec)	
16. Overall health	
17. Mental health	
18. Ability to carry out daily activities	
19. Willingness to undergo total knee replacement	
20. Weight gain between early adulthood and middle age	
21. Change in joint space width	
22. Gout	
23. Smoking	
24. Level of physical activity at work	
25. Previous contralateral knee replacement	
26. Quality of life	
27. Any past knee surgery on either knee	

Conclusion:

The ten studies that reported prediction models had different strengths and weaknesses. Most studies clearly defined their prediction models, used acceptable modelling methodologies when these were reported, reported effect estimates with reasonable precision and selected the cohort from an appropriate spectrum of patients. However, none of the studies validated their models using external data and only Riddle et al. [14] and Jiang et al. [16] addressed and used appropriate methods for missing data. Further, the methods used for predictor selection were not well described, and when they were, tended to be driven by the data available and specific interests of the researchers.

The lack of consensus on factors predictive of TKR may be due to differences in the way predictive factors were selected between the studies, the way in which missing data or possible biases in the data were treated and differences in definition of response to TKR. Minor differences in the patient cohorts and follow-up times may have also contributed somewhat to these differences. Despite this, there are no major violations of statistical methods or study design in any of the papers. Hence, there are no studies that can be simply ruled out when considering which factors to select from the literature. Instead, the review highlights the need for well-developed statistically sound prediction models for TKR and non-response to TKR that account for any missing data and possible biases that may arise in the data that is both internally and externally validated. Most importantly, none of these models were developed in our cohort and research setting of interest. That is, Australian patients aged 45 years and over attending general practice early on in their OA journey. This further highlights the need to develop two prediction models for TKR and non-response to TKR from scratch as opposed to simply

updating any of these existing models.

Appendix B. Results from predictor selection surveys issued to clinical expert panel

From Steyerberg et al. [41], the recommended approach for selecting predictors for model development includes a review of the literature as well as consultation with experts in the field under study. The literature review (summary in Appendix A) identified several candidate predictors for TKR and non-response to TKR. An adapted Delphi process was used to obtain consensus amongst experts in the field of orthopaedic surgery on these potential predictive factors and any others not identified in the literature [42]. A Delphi approach was adopted, similar to that used by Boogaard et al. [43] and Miro et al. [44].

A panel of experts was chosen based on experience and level of knowledge from the Centre for Research Excellence (CRE) for Optimising Outcomes, equity, cost effectiveness and patient Selection (OPUS) in Total Joint Replacement and consisted of orthopaedic surgeons, general practitioners, physiotherapists, epidemiologists and researchers in orthopaedics. Initially, experts were sent a link via email to the first online Google Form questionnaire [74] and were asked to indicate whether they thought the factors identified in the literature review were predictive or not predictive of the outcome. The questionnaire allowed participants to respond “no opinion” if they didn't have an opinion or were unsure. A section was included in the survey that allowed experts to draw on their knowledge and experience to identify any other factors that may be predictive of the outcomes that were not mentioned in the questionnaire. There were no restrictions on the factors that could be specified. That is, experts could identify patient clinical factors, health system factors etc. Experts were encouraged to forward the survey on to any other colleagues they considered experts in the field.

Survey responses were analysed and any new factors that were identified by the expert panel that were not listed in the first survey were included in the second survey. The second survey was resent to the panel of experts who were asked to select whether they thought these newly identified factors were predictive or not predictive of the outcomes with the option to select “no opinion.” Both surveys were open for responses over a two-week period and were sent approximately five weeks apart.

Responses to the surveys were analysed (results in Tables B 1.1 and B 1.2). The percentage of positive responses for each of the factors was calculated. Those who responded “no opinion” were excluded from the calculation for that factor. Only factors with at least 50% agreement amongst experts and routinely collected in the primary care setting (i.e. available in the EMR) were short-listed as potential predictors for use in model development.

Table B 1.1

Survey one results (N = 18)

Predictors	Responded “Yes” n (%)	Responded “No opinion” n (%)	Available in EMR Yes/No
Model one			
Age	17 (94)	–	Yes
BMI	16 (89)	–	Yes
Seeing any healthcare provider for OA	11 (73)	3 (17)	No
Knee OA grade	17 (94)	–	No
Frequency of knee pain	16 (89)	–	No
Severity of knee pain	17 (94)	–	No
Severity of knee stiffness	10 (63)	2 (11)	No
Functional difficulty	17 (94)	–	No
Use of medications for OA	9 (60)	3 (17)	Yes (prescribing only)
ROM extension	7 (50)	4 (22)	No
Knee effusion bulge sign	2 (14)	4 (22)	No
Knee flexion contracture	9 (64)	4 (22)	No
Patient initiated knee flexion	2 (15)	5 (28)	No
Maximum knee extension force divided by body weight	2 (15)	5 (28)	No
Repeated chair stand pace	4 (29)	4 (22)	No
Overall health	13 (77)	1 (6)	Yes
Mental health	12 (71)	1 (6)	Yes
Ability to carry out daily activities	15 (94)	2 (11)	No
Willingness to undergo TKR	18 (100)	–	No
Weight gain between early adulthood and middle age	7 (50)	4 (22)	Yes
Change in joint space width	13 (87)	3 (17)	No
Gout	1 (8)	5 (28)	Yes
Smoking	2 (18)	7 (39)	Yes
Level of physical activity at work	9 (56)	2 (11)	No
Previous contralateral knee replacement	14 (82)	1 (6)	Yes
Quality of life	13 (81)	2 (11)	No
Any past knee surgery on either knee	12 (71)	1 (6)	Yes
Model two			
	Responded “Yes” n (%)	Responded “No opinion” n (%)	Available in EMR and SMART Yes/No
Age	1 (6)	1 (6)	Yes
Knee OA grade	6 (35)	1 (6)	No
Number of other troublesome knees/hips	8 (47)	1 (6)	Yes
BMI	10 (59)	1 (6)	Yes
Number of comorbidities	11 [65]	1 (6)	Yes
Mental health	16 (89)	–	Yes
Presence of low back pain	9 (56)	2 (11)	No (Yes in EMR, No in SMART)
Opioid use prior to surgery	10 (63)	2 (11)	

(continued on next column)

Table B 1.1 (continued)

			Yes (Yes in EMR, No in SMART but available through linkage with PBS)
Fibromyalgia	14 (88)	2 (11)	Yes
Pain relief expectations	16 (88)	4 (22)	No

Table B 1.2

Survey two results (N = 16)

Predictors	Responded “Yes” n (%)	Responded “No opinion” n (%)	Available in EMR Yes/No
Model one			
Health insurance	10 (67)	1 (6)	No
Level of education	6 (46)	3 (19)	No
Failure of conservative treatment	15 (94)	–	No
Patient expectation that they need a TKR	15 (94)	–	No
Family or friends having good outcomes from TKR	15 (94)	–	No
Cognitive factors	11 (85)	3 (19)	No
Use of passive interventions	5 (36)	2 (13)	No
Misinformed about OA	11 (85)	3 (19)	No
Geographical location	11 (73)	1 (6)	Yes
Repeated imaging	9 (64)	2 (13)	No
Rapid worsening of knee OA in last 2 years	16 (100)	–	No
Inability to maintain relationships with family or friends due to knee OA	11 (79)	2 (13)	No
Foot position and overpronation causing medial tibial rotation	1 (7)	2 (13)	No
Health literacy	9 (60)	1 (6)	No
Model two			
Severity of comorbidities	15 (94)	–	No
Cultural background	5 (42)	4 (25)	Yes
Resilience	6 (46)	3 (19)	No
Inadequate physiotherapy pre and post knee replacement	8 (53)	1 (6)	No
Patient expectations	14 (93)	1 (6)	No
Physical inactivity	11 (85)	3 (19)	No
Patients with lower levels of pain and higher levels of function	11 (73)	1 (6)	No
Poor muscle condition	10 (67)	1 (6)	No
Axonal lumbar radiculopathy	5 (63)	8 (50)	No
Previous illicit drug use	4 (44)	7 (44)	No
Chronic pain from any source	15 (94)	–	No
Diabetes	4 (33)	4 (25)	Yes
Inconsistency between symptom and objective severity of disease	13 (93)	2 (13)	No

Appendix C. Data linkage methodology

Data linkage for model one will be performed using a privacy-protecting data linkage methodology. This involves the use of the “hashing algorithm” within the GRHANITE software system that was developed by the University of Melbourne Health and Biomedical Informatics Centre, Research Technology Unit [75]. Further information regarding the generation of patient hashes by GRHANITE is detailed by Nguyen et al. [76]. The MedicineInsight general practice data has been previously de-identified during data extraction from the respective general practice clinics, each patient has been allocated at least one hash from the GRHANITE system.

BioGrid Australia will run patient identifiers from the AOANJRR through the GRHANITE software system to create patient hashes. The MedicineInsight and AOANJRR data sets will be linked using these hashes via deterministic data linkage methods. Unique Subject Identifiers (USIs) will then be assigned to each patient and patient hashes removed.

AIHW will utilise GRHANITE to create hashes for patients within the National Death Index data set. Deterministic data linkage methods utilising the GRHANITE hashes will be used to link this data with the MedicineInsight data. Once complete, BioGrid Australia will remove any patient identifying information and hashes, and only USIs will remain to allow The University of Melbourne research team to link the three data sets: MedicineInsight, AOANJRR and NDI.

For model two, AIHW will utilise patient identifying information (first name, surname, gender, date of birth and postcode) from SMART to link this data set with PBS data using probabilistic data linkage methods. Once linkage has been performed, all patient identifying information will be removed. The final linked de-identified data set will be housed on the Sax Institute's virtual machine Secure Unified Research Environment (SURE).

Appendix D. Coding of variables

Coding of outcomes

Coding TKR from EMR and AOANJRR data (model one)

Date of TKR will be determined from the EMR data and validated against the AOANJRR [46]. Similarly, dates of death will be validated against the NDI [47]. Any incomplete or inaccurate data in the EMR relating to these variables will be corrected prior to model development. Those who underwent their first primary TKR (on a particular knee) within the study period will be coded as 1 = “Underwent TKR”. Patients who did not undergo TKR will be censored and coded as 0 = “Did not undergo TKR”, and those who died as 2 = “Died”. Given that the AOANJRR captures all TKRs performed in Australia, there will be no loss to follow-up in this study as outcomes during the study period are known for all patients within our cohort (excluding those who died).

Coding non-response to TKR from SMART (model two)

The Outcome Measures in Rheumatology-Osteoarthritis Research Society International (OMERACT-OARSI) responder criteria [8,40] will be used to classify patients as either responders (0 = “responder to TKR”) or non-responders (1 = “non-responder to TKR”). To code non-response to TKR, the three subscales of the WOMAC score: pain, stiffness and function will be first normalised. Changes in the WOMAC subscale scores between 12 months (post-surgery) and baseline will be calculated as per the OMERACT-OARSI responder criteria.

Coding of predictors

Coding of BMI (models one and two)

BMI will be coded as a continuous variable in both models and all measurements will be in kg/m². For model one, BMI will be coded from the observations text field within the EMR data. Only BMI entries recorded within the year prior to baseline will be included. Where there are two BMI observations recorded on the same day, the average of the two recordings will be used. For model two, BMI prior to surgery will be coded from the numeric BMI data field within SMART. Patients with implausible values for BMI will be coded as having missing data (see Data Inspection for how missing data will be addressed).

Coding use of medication/s for OA (model one)

Prescriptions data from the EMR will be used to determine whether patients are using OA medications at baseline. The osteoarthritis medications that will be considered are listed by Anatomical Therapeutic Chemical (ATC) class in Table D 1.1 and were identified by clinical researchers in this study. A full list of medications by ATC code are provided in Table D 1.2.

Table D 1.1
Medications used for osteoarthritis management by ATC class

ATC class	ATC class name
H02	Corticosteroids for systemic use
M01	Anti-inflammatory and antirheumatic products
M02	Topical products for joint and muscular pain
M09	Other drugs for disorders of the musculo-skeletal system
N01	Anaesthetics
N02	Analgesics
N06	Psychoanalytics

Table D 1.2
Osteoarthritis medication by ATC code

ATC code	Drug name
H02AB01	betamethasone
H02AB02	dexamethasone
H02AB04	methylprednisolone
H02AB06	prednisolone
H02AB07	prednisone
H02AB08	triamcinolone
H02AB09	hydrocortisone
H02AB10	cortisone
H02AB15	meprednisone
H02BX01	methylprednisolone, combinations
M01AB01	indometacin
M01AB02	sulindac
M01AB05	diclofenac
M01AB15	ketorolac
M01AB51	indometacin, combinations
M01AB55	diclofenac, combinations
M01AC01	piroxicam
M01AC06	meloxicam
M01AC56	meloxicam, combinations
M01AE01	ibuprofen
M01AE02	Naproxen
M01AE03	ketoprofen
M01AE14	dexibuprofen
M01AE17	dexketoprofen
M01AE51	ibuprofen, combinations
M01AE52	naproxen and esomeprazole
M01AE53	ketoprofen, combinations
M01AE56	naproxen and misoprostol
M01AG01	mefenamic acid
M01AH01	Celecoxib
M01AH02	Rofecoxib
M01AH04	Parecoxib
M01AH05	Etoricoxib
M01AH05	Glucosamine
M01AX25	chondroitin sulfate
M01BA03	acetylsalicylic acid and corticosteroids

(continued on next page)

Table D 1.2 (continued)

ATC code	Drug name
M02AA07	Piroxicam
M02AA10	ketoprofen
M02AA12	Naproxen
M02AA13	Ibuprofen
M02AA15	diclofenac
M02AA23	indometacin
M02AA27	dexketoprofen
M02AA28	piketoprofen
M02AB	capsaicin and similar agents
M02AB01	Capsaicin
M02AB02	zucapsaicin
M09AX01	hyaluronic acid
N01BX04	Capsaicin
N02AA01	Morphine
N02AA03	hydromorphone
N02AA05	oxycodone
N02AA08	dihydrocodeine
N02AA51	morphine, combinations
N02AA53	hydromorphone and naloxone
N02AA55	oxycodone and naloxone
N02AA56	oxycodone and naltrexone
N02AA58	dihydrocodeine, combinations
N02AA59	codeine, combinations excl. psycholeptics
N02AA79	codeine, combinations with psycholeptics
N02AB02	Pethidine
N02AB03	Fentanyl
N02AB52	pethidine, combinations excl. psycholeptics
N02AB72	pethidine, combinations with psycholeptics
N02AC04	dextropropoxyphene
N02AC52	methadone, combinations excl. psycholeptics
N02AC54	dextropropoxyphene, combinations excl. psycholeptics
N02AC74	dextropropoxyphene, combinations with psycholeptics
N02AE01	buprenorphine
N02AG01	morphine and antispasmodics
N02AG03	pethidine and antispasmodics
N02AG04	hydromorphone and antispasmodics
N02AJ01	dihydrocodeine and paracetamol
N02AJ02	dihydrocodeine and acetylsalicylic acid
N02AJ03	dihydrocodeine and other non-opioid analgesics
N02AJ06	codeine and paracetamol
N02AJ07	codeine and acetylsalicylic acid
N02AJ08	codeine and ibuprofen
N02AJ09	codeine and other non-opioid analgesics
N02AJ13	tramadol and paracetamol
N02AJ14	tramadol and dexketoprofen
N02AJ15	tramadol and other non-opioid analgesics
N02AJ17	oxycodone and paracetamol
N02AJ18	oxycodone and acetylsalicylic acid
N02AJ19	oxycodone and ibuprofen
N02AX02	Tramadol
N02AX06	tapentadol
N02BA01	acetylsalicylic acid
N02BE01	Paracetamol
N06AX21	Duloxetine

The drug dosage, strength, frequency, number of repeats and medication quantity will be used to estimate the prescription end date and hence whether the medication was in use at baseline. Only prescriptions recorded or issued in the 12 months prior to baseline will be included given that prescriptions can only last for 12 months in Australia. For prescriptions with a range of dosages (e.g. 1–2 tablets daily), the minimum dosage and hence maximum end date will be assumed. Medications that have been prescribed “as needed” (pro-re nata) will have missing estimated end dates as dosages are unknown. Prescriptions with daily dosages above that recommended in the Australian Medicines Handbook [77] will be treated as implausible dosages and hence estimated end dates will be missing for these. Similarly, prescriptions marked as “Pharmaceutical Benefits Scheme” (PBS) or “Repatriation Pharmaceutical Benefits Scheme” (rPBS) that state medication quantities or number of repeats greater than that specified on the Australian Government Department of Health PBS and rPBS website will also be coded as missing for estimated prescription end date due to implausible data. Patients will be assumed to have missing data for the use of OA medications at baseline if they have prescriptions issued in the year before baseline with missing estimated end dates, unless they have been prescribed another OA medication in which an end date can be estimated that is beyond baseline. The coding methodology for OA prescriptions is summarised in Figure D.1.1.

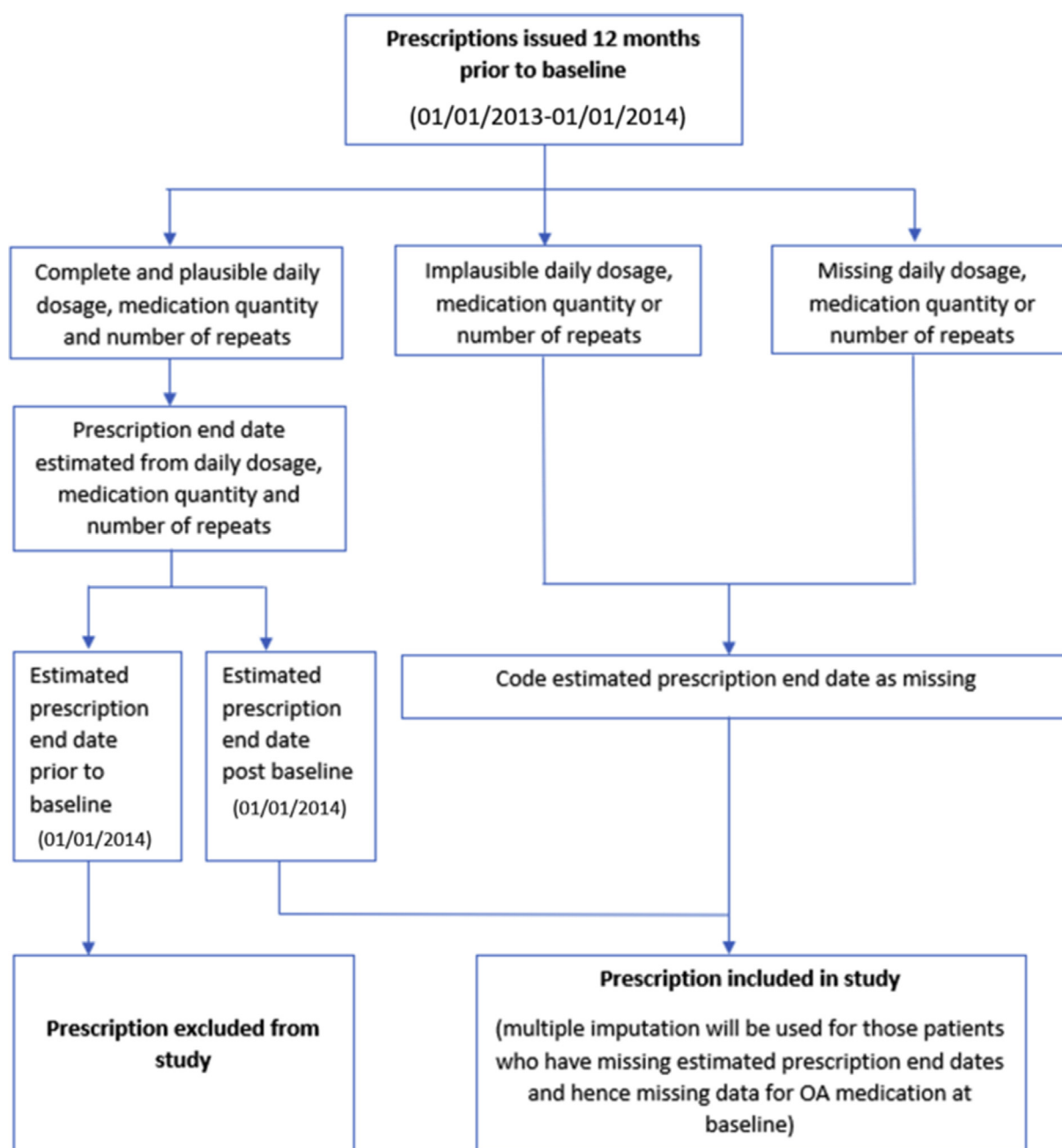


Fig. D 1.1. Coding methodology for OA medications

Coding use of opioids prior to surgery (model two)

Opioid use at baseline will be coded from linked PBS data using ATC class code N02A. All other medications listed in Table D 1.2 will be coded as non-opioid OA medications. Only opioid and non-opioid OA prescriptions dispensed in the 12 months prior to the patient's surgery date will be included. This variable will be coded as an ordinal categorical variable with the following levels: 0 = "no OA medication use", 1 = "use of non-opioid medication only", 2 = "use of opioid medication only" and 3 = "use of both non-opioid and opioid medication."

Multimorbidity (models one and two)

Table D 1.3 lists the chronic conditions from the CCI and the BEACH study that will be coded. Chronic problems included in the CCI and BEACH study that are not clearly recorded in general practice EMRs (localised solid tumour, shoulder syndrome, chronic skin ulcer, chronic acne, vertiginous syndrome and neck syndrome) will be excluded here to avoid bias multimorbidity counts.

Table D 1.3

Chronic conditions used for multimorbidity count

Charlson Comorbidity Index (CCI)	Chronic conditions most frequently managed in primary care from BEACH study
Myocardial infarction (MI)	Hypertension
Congestive heart failure (CHF)	Diabetes mellitus
Peripheral vascular disease (PVD)	Lipid disorder
Cerebrovascular accident (CVA)/transient ischaemic attack (TIA)	Dyspepsia
Dementia	Asthma
Chronic obstructive pulmonary disease (COPD)	Atrial fibrillation (AF)
Connective tissue disease	Malignant neoplasms of the skin
Peptic ulcer disease	Osteoporosis
Liver disease	Hypothyroidism/myxoedema
Diabetes mellitus	Ischaemic heart disease (IHD)
Hemiplegia	Chronic obstructive pulmonary disease (COPD)
Chronic kidney disease (CKD)	Gout
Leukaemia	Migraine
Lymphoma	Congestive heart failure (CHF)
Malignant tumour	Chronic pain not otherwise specified (model one only)
Human immunodeficiency virus (HIV)/Acquired immunodeficiency syndrome (AIDs)	Dementia
	Chronic kidney disease (CKD)

Mental health conditions will not be included in multimorbidity count as these will be included as separate predictors in both models. Chronic pain not otherwise specified will be included in the multimorbidity count for model one only as these conditions are not well documented in SMART.

For model one, the diagnoses text field in the EMR will be searched for each of the chronic conditions listed in Table D 1.3 using the free text terms listed in Table D 1.4. These terms were defined by NPS MedicineWise [78] and by a general practitioner who is a researcher on this study. All variations of these terms will be searched, including possible spelling errors. Any free text with negative terms or uncertain diagnoses will not be included (i.e. “negative”, “unconfirmed”, “no”, “probable”, “likely”, “suspected”, “borderline”, “screening”, “possible”, “fear of”, “high risk”, “family history”, “survived” etc.) Each of the diagnoses will be coded in binary format (1 = “has condition”, 0 = “does not have condition”). Given that GPs may record possible diagnoses in the EMR that are yet to be proven, patients with diagnoses that have been recorded with uncertainty or that are non-specific will be treated as not having the condition unless a later entry confirms the diagnosis. For each patient, multimorbidity count at baseline will be calculated by summing the number of multimorbidities with onset diagnosis date prior to baseline. Multimorbidity count will be set to missing for patients with one or more missing diagnoses onset dates.

Table D 1.4

Free text terms for chronic conditions

Condition	Free text terms
Myocardial infarction (mi)	Mi Ami Acute ischaemic heart disease Myocardial infarction Heart attack Stemi Subendocardial infarct Angiogram-mild heart disease
Peripheral vascular disease	Pvd Peripheral vascular disease Diabetes vascular disease Arteritis-diabetes mellitus Buerger's disease Obliterative vascular disease Diabetic endarteritis Thrombngitis obliterans Arteriosclerosis obliterans Diabetes with vascular changes Occlusive vascular disease Peripheral arterial disease
Cva/tia	Cerebral haemorrhage Cerebral infarction Cerebrovascular accident Cva Haemorrhage-intracerebral Stroke Intracerebral bleed Intracranial haemorrhage Lacunar infarct Migrainous stroke
Connective tissue disease	Rheumatoid arthritis Polymyalgia rheumatica Ra Sle Scleroderma Lupus Sle Systemic lupus erythematosus

(continued on next column)

Table D 1.4 (continued)

Condition	Free text terms
Peptic ulcer disease	Polymyositis Dermatomyositis Peptic ulcer Ulcerative reflux disease Gastric ulcer Stomach ulcer Gord with ulceration
Liver disease	Chronic hepatitis Cirrhosis Fatty liver Liver failure Liver disease Alcoholic liver Liver dysfunction Liver damage Nash Nafld Non-specific hepatitis Hepatitis- collagen disease Cholestatic hepatitis Autoimmune hepatitis Drug induced hepatitis
Hemiplegia	Hemiplegia
Leukaemia	Leukaemia
Lymphoma	Lymphoma
Malignant tumour	Cancer Malig Carcino Metast (Note: skin cancer terms excluded, pre-cancer terms excluded, negative cancer markers excluded, malignant hyperthermia excluded, lentigo maligna excluded)
Hiv/aids	Hiv Aids Immunodeficiency virus
Hypertension	Hypertension HT High blood pressure High bp H/t Hypertensive Htn Bp high Labile bp Bp labile Labile blood pressure Blood pressure labile Hbp Elevated blood pressure Raised blood pressure Antihypertensive agent
Diabetes mellitus	Hbpm T2dm Diab Niddm Mellitus Ketoac Dka Osmolar Insulin
Lipid disorder	Hyperlip Hyperchol Hypertrig High chol Hyperlipoprotein High lipid Dyslip
Dyspepsia/oesophageal disease	Gor Hh Heartburn Dyspepsia Reflux Barrett Indigest Petic ulcer Hyperacidity Sliding hernia

(continued on next column)

Table D 1.4 (continued)

Condition	Free text terms
Asthma	Regurgitation
	Oesophagitis
	Hiatus hernia
	Epigastric pain
	Belching
	Helicobacter pylori
	Asthma
	Samter
Atrial fibrillation	Wheezy bronchitis
	Asthmaticus
	Asthmoid
	Af
	A/f
Malignant neoplasms of the skin	Atrial fib
	A.f.
	A FIB
	Sec
	BCC
	Squamous cell
Osteoporosis	Basal cell
	Melanom
	Osteoporosis
Hypothyroidism/myxoedema	Osteoporotic
	Hypothyroid
	Myxoedema
	Hashimotos
Ischaemic heart disease (ihd)	Thyroiditis
	Ami
	MI
	Acute myocardial infarction
	Myocardial infarction
	Heart attack
	Acute ischaemic heart disease
	Ischaemic heart disease
	Angiogram-mild heart disease
	Subendocardial infarct
	Subendocardial myocardial infarct
	Stemi
	Nstemi
	Angina
	Cad
	Ihd
	Cabg
	Angioplasty
	Atherosclerotic heart disease
	Coronary heart disease
	Heart disease
	Preinfarction syndrome
	Pta
	Coronary insufficiency
	Coronary occlusion
	Coronary artery block
	Ischaemic vascular disease
	Occlusion, coronary
	Occlusion – coronary
	Obstruct & stent & aorta
	Obstruct & bypass & biliary
	Cardiac arrest
Chronic obstructive pulmonary disease (copd)	Copd
	Coad
	Emphysema
	Emphysematous
	Chronic airway/s limitation
	Chronic bronchitis
	Chronic obstructive airway
	Chronic obstructive pulmonary
	Bronchitis- chronic
	Chronic airways disease
	Chronic airflow limitation
	Chronic airway/s infection
	Cal
	Bronchitis chronic
Gout	Gout
Migraine	Migraines
Congestive heart failure	Migratory
	Heart failure

(continued on next column)

Table D 1.4 (continued)

Condition	Free text terms
Chronic pain	Cardiac failure
	Ccf
	Chf
	Cor pulmonae
	Lhf
	Rhf
	Lvf
	Rvf
	Cardiac dysfunction
	Cardiomyopathy
	Ventricular failure
	Ventricular diastolic dysfunction
	Headache
	Plantar fascitis
	Carpal tunnel
	Rheumatoid arthritis
	Dysarthrosis
	Shingles
	Neuralgia
	Compressed nerve
	Spondylosis
	Back pain
	Back problem
	Neck pain
	Neck problem
	Thoracic pain
	Thoracic problem
DEMENTIA	Sciatica
	Bulging disc
	Dementia
	Demented state
	Dementing illness
CHRONIC KIDNEY DISEASE	Dementio vascular
	Alzheimers
	Chronic kidney disease
	Chronic renal disease
	Ckd
	Renal impairment
	Dialysis
	Kidney disease
	Kidney impairment
	Kidney damage
	Kidney failure
	Kidney insufficiency
	Kidney end stage
	Renal disease
	Renal impairment
	Renal damage
	Renal failure
	Renal insufficiency
	Renal end stage
	Uraemia
	Capd
	Crf
	Peritoneal catheter

A similar approach will be used for model two, however, diagnoses will be coded from the SMART data set instead.

Coding mental health condition (models one and two)

The diagnoses text field within the EMR (model one) will be searched for the mental health conditions listed in Table D 1.5. The diagnosis onset date will be used to determine whether the patient has a mental health condition at baseline. Mental health condition will be coded in binary format: 1 = “diagnosis of mental health condition” and 0 = “no diagnosis of mental health condition”. Patients with a missing diagnosis onset date for a mental health condition will be coded as missing, unless they have other mental health conditions recorded with diagnoses onset dates that precede baseline.

Table D 1.5
Free text terms for mental health conditions

Condition	Free text terms
Depression	Depression DEPRESS MOOD DEPRESSED DEPRESSIVE MELANCHOL
Anxiety	Gad ANXIETY ANXIOUS ANXIOLYTIC ANTI-ANXIETY AGENT PRESCRIPTION GENERALISED ANXIETY DISORDER
Post-traumatic stress disorder	Post traumatic stress disorder POST-TRAUMATIC PTSD
Obsessive compulsive disorder	Obsessive compulsive disorder OBSESSIVE OCD
Anorexia	Anorexia
Bulimia	Bulimia
Bi-polar	Bipolar BI-POLAR
Schizophrenia	Schizophrenia ADJUSTMENT DISORDER
Dissociative disorders and other	Acute stress NEUROTIC DEPRESSION PHOBIA PANIC NERVOUS BREAKDOWN STRESS DISORDER PSYCHOGENIC AGORAPHOBIA IRRATIONAL FEAR PERSONALITY DISORDER

The pre-operation comorbidities text field within SMART will be searched for the conditions listed in Table D 1.5 to code mental health conditions for model two.

Coding weight gain between early adulthood and middle age (model one)

The weight text field within the EMR will be cleaned such that any non-numeric characters are removed and all measurements are in kilograms. Weight gain between early adulthood and middle age will be calculated by subtracting weight measurements between the ages of 18–21 years from weight at 45–65 years [20]. Those who experience weight loss will be coded as having zero weight gain and weight gain will be coded as a continuous variable. Weight gain for patients with missing weight measurements at the ages of 18–21, or 45–65, or both, will be coded as missing.

Coding previous/contralateral knee replacement (model one)

Date and side of TKR from the diagnosis text field in the EMR will be used to determine whether patients underwent TKR prior to baseline. This will be validated using data from the AOANJRR. Given that this study focuses on time to primary TKR, patients will be coded as having had a contralateral knee replacement if they underwent TKR during the study period and there is a recording of TKR on the opposite knee prior to baseline. Patients who do not undergo TKR during the study period but who have undergone a TKR prior to baseline will also be coded as having had a previous knee replacement. This variable will be coded as binary: 1 = “underwent contralateral/previous TKR” and 0 = “did not undergo contralateral/previous TKR.”

Coding any past knee surgery (model one)

The diagnosis field within the EMR will be searched for the knee surgeries listed in Table D 1.6 to identify patients who underwent knee surgeries other than TKR prior to baseline. This list was compiled by researchers in this study who are experts in joint replacement. The variable will be binary with 1 = “past knee surgery” and 0 = “no past knee surgery.” Those with recordings of past knee surgeries (other than TKR) in their EMR but with missing procedure dates will be coded as missing for this predictor.

Table D 1.6
Past knee surgeries (other than TKR)

Arthroscopy	Meniscus repair
Open reduction knee	Meniscectomy
Open repair knee	Lateral release
Knee reconstruction	Anterior cruciate ligament repair (ACL repair)
Cruciate ligament repair	Medial collateral ligament repair (MCL repair)
Clean out knee	Osteotomy knee
Debridement knee	Knee chondroplasty
Fracture tibial plateau with screws	Fracture tibial plateau with repair
Supracondylar fracture femur pin	Avulsion fracture femoral condyle
Periprosthetic fracture femoral condyle	Arthrotomy

Coding geographical location (model one)

Patient geographical location is based on the patient's residential postcode recorded within the EMR and is determined according to the Australian Bureau of Statistics (ABS) Australian Statistical Geography Standard (ASGS) remoteness areas [79] which divides states and territories according to access to services: major cities, inner regional, outer regional, remote and very remote. This predictor will be coded as a categorical variable: 0 = "major cities", 1 = "inner regional", 2 = "outer regional", 3 = "remote and very remote."

Fibromyalgia (model two)

The pre-operation comorbidities text field within SMART will be searched for "fibromyalgia." This predictor will be coded in binary format: 1 = "diagnosis of fibromyalgia" and 0 = "no diagnosis of fibromyalgia."

Appendix E. Optimism corrected discrimination and regression

The process for obtaining optimism corrected discrimination will be as follows [41,66]:

1. Fit the model corresponding to a threshold frequency of 60% to an imputed data set and calculate the C statistic within that data set (apparent performance).
2. Apply the same model to a bootstrapped sample of the first imputed data set and calculate the C statistic within this bootstrapped data set.
3. Calculate the optimism as the difference between the C statistic from the bootstrapped sample and the first imputed sample.
4. Repeat steps 1–3 approximately 100 times to get a stable estimate of the optimism in this imputed data set.
5. Repeat steps 1–4 for each of the 30 imputed data sets to get stable estimates of optimism for each imputed data set.
6. Apply the optimism correction to the apparent performance for each imputed data set to get the optimism corrected performance of each imputed data set.
7. Average the optimism corrected performance across all imputed data sets to get an estimate of the overall optimism corrected performance for the model with 60% threshold frequency.
8. Repeat steps 1–7 for the remaining threshold frequencies (70%, 80% and 90%).

The process for obtaining calibration slopes and hence optimism corrected regression estimates will be as follows [41,66]:

1. Fit the model corresponding to a threshold frequency of 60% in a bootstrapped sample of the first imputed data set.
2. Use the regression coefficients in this bootstrapped sample to calculate the linear predictor for each patient in the original first imputed data set (i.e. use the values of predictors in this first imputed data set).
3. Use the outcomes of patients in the first imputed data set and the calculated linear predictors to obtain an estimate of the slope of the linear predictor.
4. Repeat steps 1–3 for each of the 100 bootstrapped samples from the first imputed data set to obtain a stable estimate of the mean of the slope of the linear predictor for the first imputed data set. This is the estimated shrinkage factor for the first imputed data set.
5. Repeat steps 1–4 for each imputed data set.
6. Average the estimated shrinkage factors across all imputed data sets to obtain the overall shrinkage factor.
7. Apply the overall shrinkage factor to the regression estimates for the final model with threshold frequency 60%.
8. Repeat steps 1–7 for each remaining threshold frequency (70%, 80% and 90%).

References

- [1] Australian Institute of Health and Welfare, Osteoarthritis Snapshot, what is Osteoarthritis? - Australian Institute of Health and Welfare, 2018. Available from: <https://www.aihw.gov.au/reports/chronic-musculoskeletal-conditions/osteoarthritis/contents/what-is-osteoarthritis>.
- [2] I.N. Ackerman, C. Pratt, A. Gorelik, D. Liew, Projected burden of osteoarthritis and rheumatoid arthritis in Australia: a population-level analysis, *Arthritis Care Res. (Hoboken)* 70 (6) (2018 Jun) 877–883, <https://doi.org/10.1002/acr.23414>. Epub 2018 Apr 12. PMID: 28898565.
- [3] Australian Institute of Health and Welfare, Osteoarthritis, Australian Institute of Health and Welfare, 2017 [Internet]. AIHW website. Available from: <https://www.aihw.gov.au/reports/arthritis-other-musculoskeletal-conditions/osteoarthritis/contents/what-is-osteoarthritis>.
- [4] L. March, H. Bagga, Epidemiology of osteoarthritis in Australia, *Med. J. Aust.* 180 (2004) S6–S10. Available from: <https://search-proquest-com.ezp.lib.unimelb.edu.au/docview/1802213385/fulltextPDF/B1BF93C5A6C4B77PQ/1?accountid=12372>.
- [5] C. Steiner, R. Andrews, M. Barrett, M. Weiss, HCUP Projections Mobility/Orthopedic Procedures 2003 to 2012 Report # 2012-03, 2003. Available from: <http://www.hcup-us.ahrq.gov/reports/projections/2012-03.pdf>.
- [6] T.N. Peel, A.C. Cheng, D. Liew, K.L. Buising, J. Lisik, K.A. Carroll, et al., Direct hospital cost determinants following hip and knee arthroplasty, *Arthritis Care Res. (Hoboken)* 67 (6) (2015) 782–790. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=25470687>.
- [7] K. Alzahrani, R. Gandhi, J. Debeer, D. Petrucci, N. Mahomed, Prevalence of clinically significant improvement following total knee replacement, *J. Rheumatol.* 38 (4) (2011 Apr 1) 753–759. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21239743>.
- [8] M.M. Dowsey, T. Spelman, P.F.M. Choong, Development of a prognostic nomogram for predicting the probability of nonresponse to total knee arthroplasty 1 Year after surgery, *J. Arthroplasty* 31 (8) (2016 Aug 1) 1654–1660. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0883540316001212>.
- [9] R.S. Hinman, P.J.A. Nicolson, F.L. Dobson, K.L. Bennell, Use of nondrug, nonoperative interventions by community-dwelling people with hip and knee osteoarthritis, *Arthritis Care Res. (Hoboken)* 67 (2) (2015 Feb 1) 305–309. Available from: <http://pubmed.ncbi.nlm.nih.gov/25048646/>.
- [10] J. Chua, A.M. Briggs, P. Hansen, C. Chapple, J.H. Abbott, Choosing interventions for hip or knee osteoarthritis - what matters to stakeholders? A mixed-methods study, *Osteoarthritis Cartilage* 28 (2020 Apr 9) 100062.
- [11] J.R. Lewis, S.S. Dhaliwal, K. Zhu, R.L. Prince, A predictive model for knee joint replacement in older women, *PLoS One* 8 (12) (2013), e83665 [Electronic Resour [Internet] Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=24349541>].
- [12] J.A. Zeni Jr., M.J. Axe, L. Snyder-Mackler, J.A. Zeni Jr., M.J. Axe, L. Snyder-Mackler, Clinical predictors of elective total joint replacement in persons with end-stage knee osteoarthritis, *BMC Musculoskel. Disord.* 11 (86) (2010) 86. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med12&AN=50904709>.
- [13] G.A. Hawker, J. Guan, R. Croxford, P.C. Coyte, R.H. Glazier, B.J. Harvey, et al., A prospective population-based study of the predictors of undergoing total joint arthroplasty, *Arthritis Rheum.* 54 (10) (2006 Oct) 3212–3220. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17009255>.
- [14] D.L. Riddle, X. Kong, W.A. Jiranek, Factors associated with rapid progression to knee arthroplasty: complete analysis of three-year data from the osteoarthritis

- initiative, Joint, Bone, Spine Rev du Rhum 79 (3) (2012) 298–303. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=21727020>.
- [15] G.A. Hawker, E.M. Badley, C.M. Borkhoff, R. Croxford, A.M. Davis, S. Dunn, et al., Which patients are most likely to benefit from total joint arthroplasty? *Arthritis Rheum.* 65 (5) (2013) 1243–1252. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=23459843>.
 - [16] Y. Jiang, M.T. Sanchez-Santos, A.D. Judge, D.W. Murray, N.K. Arden, Predictors of patient-reported pain and functional outcomes over 10 Years after primary total knee arthroplasty: a prospective cohort study, *J. Arthroplasty* 32 (1) (2017) 92–100, e2. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=27444848>.
 - [17] H.R.B.A. Razak, C.S. Tan, Y.J.D. Chen, H.N. Pang, K.J. Darren Tay, P.L. Chin, et al., Age and preoperative knee society score are significant predictors of outcomes among asians following total knee arthroplasty, *J. Bone Jt. Surg. Am. Vol.* 98 (9) (2016) 735–741. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=27147686>.
 - [18] E. Lungu, F. Desmeules, C.E. Dionne, É.L. Belzile, P.-A. Vendittoli, Prediction of poor outcomes six months following total knee arthroplasty in patients awaiting surgery, *BMC Musculoskel. Disord.* 15 (1) (2014 Dec 8) 299. Available from: <http://bmcmusculoskeletaldisord.biomedcentral.com/articles/10.1186/1471-2474-15-299>.
 - [19] N.D. Clement, P.J. Jenkins, D. MacDonald, Y.X. Nie, J.T. Patton, S.J. Breusch, et al., Socioeconomic status affects the Oxford knee score and short-form 12 score following total knee replacement, *Bone Joint Lett.* J 95-B (1) (2013) 52–58. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=223307673>.
 - [20] Y. Wang, A.E. Wluka, J.A. Simpson, G.G. Giles, S.E. Graves, R.N. de Steiger, et al., Body weight at early and middle adulthood, weight gain and persistent overweight from early adulthood are predictors of the risk of total knee and hip replacement for osteoarthritis, *Rheumatology* 52 (6) (2013) 1033–1041. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=23362222>.
 - [21] O. Bruyere, C. Cooper, K. Pavelka, V. Rabenda, F. Buckinx, C. Beaudart, et al., Changes in structure and symptoms in knee osteoarthritis and prediction of future knee replacement over 8 years, *Calcif. Tissue Int.* 93 (6) (2013) 502–507. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=23995765>.
 - [22] E. Ageberg, G. Engstrom, M. Gerhardsson de Verdier, J. Rollof, E.M. Roos, L.S. Lohmander, Effect of leisure time physical activity on severe knee or hip osteoarthritis leading to total joint replacement: a population-based prospective cohort study, *BMC Musculoskel. Disord.* 13 (2012) 73. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=22595023>.
 - [23] M. Dougados, M.A. D'Agostino, M. Le Bars, G. Baron, N. Schmidely, R. Wakefield, et al., Clinical and ultrasonographic predictors of joint replacement for knee osteoarthritis: results from a large, 3-year, prospective EULAR study, *Ann. Rheum. Dis.* 1 (4) (2009) S237. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med12&AN=70189647>.
 - [24] K. Magnusson, K. Scurrah, R. Orstavik, T. Nilsen, O. Furnes, K. Hagen, Physical activity, body weight and risk of hip and knee arthroplasty due to osteoarthritis: causal inference from the nor-twin OA study, *Osteoarthritis Cartil* 25 (Suppl: S179–80) (2017). Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=620926855>.
 - [25] A.S. Nicholls, A. Kiran, M.K. Javadi, D.J. Hart, T.D. Spector, A.J. Carr, et al., Change in body mass index during middle age affects risk of total knee arthroplasty due to osteoarthritis: a 19-year prospective study of 1003 women, *Knee* 19 (4) (2012) 316–319. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=21782451>.
 - [26] G.G. Teng, Y.Y. Leung, L.W. Ang, J.M. Yuan, W.P. Koh, Gout and risk of knee replacement for severe knee osteoarthritis in the Singapore Chinese Health Study, *Osteoarthritis Cartil* 25 (12) (2017), 1962–8. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=28757187>.
 - [27] K.M. Leyland, A. Judge, M.K. Javadi, A. Diez-Perez, A. Carr, C. Cooper, et al., Obesity and the relative risk of knee replacement surgery in patients with knee osteoarthritis: a prospective cohort study, *Arthritis Rheum.* 68 (4) (2016) 817–825. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=26556722>.
 - [28] M.C. Hochberg, K. Favors, J.D. Sorkin, Quality of life and radiographic severity of knee osteoarthritis predict total knee arthroplasty: data from the osteoarthritis initiative, *Osteoarthritis Cartil* 1 (2013) S11. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med15&AN=71464696>.
 - [29] Y.Y. Leung, L.W. Ang, J. Thumboo, R. Wang, J.M. Yuan, W.P. Koh, Cigarette smoking and risk of total knee replacement for severe osteoarthritis among Chinese in Singapore—the Singapore Chinese health study, *Osteoarthritis Cartil* 22 (6) (2014) 764–770. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=24680935>.
 - [30] G. Mnatzaganian, P. Ryan, C.M. Reid, D.C. Davidson, J.E. Hiller, Smoking and primary total hip or knee replacement due to osteoarthritis in 54,288 elderly men and women, *BMC Musculoskel. Disord.* 14 (2013) 262. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=24006845>.
 - [31] H. Apold, H.E. Meyer, L. Nordsletten, O. Furnes, V. Baste, G.B. Flugsrud, Risk factors for knee replacement due to primary osteoarthritis, a population based, prospective cohort study of 315,495 individuals, *BMC Musculoskel. Disord.* 15 (2014) 217. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=24957045>.
 - [32] S.R. Smith, J. Bido, J.E. Collins, H.Y. Yang, J.N. Katz, E. Losina, et al., Impact of preoperative opioid use on total knee arthroplasty outcomes, *J Bone Jt Surg - Am* 99 (10) (2017) 803–808. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=616353693>.
 - [33] C.M. Brummett, A.G. Urquhart, A.L. Hassett, A. Tsodikov, B.R. Hallstrom, N.I. Wood, et al., Characteristics of fibromyalgia independently predict poorer long-term analgesic outcomes following total knee and hip arthroplasty, *Arthritis Rheum.* 67 (5) (2015) 1386–1394. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=25772388>.
 - [34] M. Gonzalez Saenz de Tejada, A. Escobar, A. Bilbao, C. Herrera-Espineira, L. Garcia-Perez, F. Aizpuru, et al., A prospective study of the association of patient expectations with changes in health-related quality of life outcomes, following total joint replacement, *BMC Musculoskel. Disord.* 15 (2014) 248. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=25055728>.
 - [35] C. Tilbury, M.J. Holtslag, R.L. Tordoir, C.S. Leichtenberg, S.H.M. Verdegaaal, H.M. Kroon, et al., Outcome of total hip arthroplasty, but not of total knee arthroplasty, is related to the preoperative radiographic severity of osteoarthritis, *Acta Orthop.* 87 (1) (2016) 67–71. Available from: <https://search.ebscohost.com/login.aspx?direct=true&db=ccm&AN=112062549&site=ehost-live>.
 - [36] A. Utrillas-Compaired, B.J. De la Torre-Escudé, A.J. Tebar-Martínez, A. Asunsolo-Del Barco, Does preoperative psychologic distress influence pain, function, and quality of life after TKA? *Clin. Orthop. Relat. Res.* 472 (8) (2014) 2457–2465. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=24671514>.
 - [37] L.A. Liljensoe, J.O. Lauersen, K. Soballe, I. Mechlenburg, Overweight preoperatively impairs clinical outcome after knee arthroplasty: a cohort study of 197 patients 3-5 years after surgery, *Acta Orthop.* 84 (4) (2013) 392–397. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=23992141>.
 - [38] D.P. Williams, C.M. Blakey, S.G. Hadfield, D.W. Murray, A.J. Price, R.E. Field, Long-term trends in the Oxford knee score following total knee replacement, *Bone Joint Lett.* J 95-B (1) (2013) 45–51. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med7&AN=23307672>.
 - [39] J.E. Collins, L.A. Donnell-Fink, H.Y. Yang, I.M. Usiskin, E.C. Lape, J. Wright, et al., Effect of obesity on pain and functional recovery following total knee arthroplasty, *Eff Obes Pain Funct Recover Follow Total Knee Arthroplast Bone Jt Surg - Am* 99 (21) (2017) 1812–1818. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med1&AN=29088035>.
 - [40] A. Escobar, M. Gonzalez, J.M. Quintana, K. Vrotsou, A. Bilbao, C. Herrera-Espineira, et al., Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement. Identification of cut-off values, *Osteoarthritis Cartil* 20 (2) (2012 Feb 1) 87–92. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22155074>.
 - [41] E.W. Steyerberg, *Clinical Prediction Models: a Practical Approach to Development, Validation, and Updating*, Springer, 2009, p. 497.
 - [42] S. Thangaratnam, C.W. Redman, The Delphi technique, *Obstet. Gynaecol.* 7 (2) (2005 Apr 1) 120–125, <https://doi.org/10.1576/toag.7.2.120.27071>. Available from: <https://doi.org/10.1576/toag.7.2.120.27071>.
 - [43] S. Boogaard, M.W. Heymans, J. Patijn, H.C. de Vet, C.G. Faber, M.L. Peters, et al., Predictors for persistent neuropathic pain—a Delphi survey, *Pain Physician* 14 (6) (2011), 559–68. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22086097>.
 - [44] J. Miró, R. Nieto, A. Huguet, Predictive factors of chronic pain and disability in whiplash: a Delphi poll, *Eur. J. Pain* 12 (1) (2008 Jan 1) 30–47, <https://doi.org/10.1016/j.ejpain.2007.02.004>. Available from: <https://doi.org/10.1016/j.ejpain.2007.02.004>.
 - [45] N. MedicineWise, *MedicineInsight Data Book*, 2018. Available from: www.medicinesinsight.org.au.
 - [46] South Australian Health and Medical Research Institute, Home - AOANJR [Internet], 2019. Available from: <https://aoanjrh.sahmri.com/>.
 - [47] Welfare AI of H and National Death Index (NDI), Australian Institute of Health and Welfare, 2020. Available from: <https://www.aihw.gov.au/about-our-data/our-data-collections/national-death-index>.
 - [48] K.G.M. Moons, P. Royston, Y. Vergouwe, D.E. Grobbee, D.G. Altman, Prognosis and prognostic research: what, why, and how? *BMJ* 338 (2009 Feb 23) b375. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19237405>.

- [49] M.C. Inacio, N.L. Pratt, E.E. Roughead, S.E. Graves, Comparing co-morbidities in total joint arthroplasty patients using the RxRisk-V, Elixhauser, and Charlson Measures: a cross-sectional evaluation, *BMC Musculoskel. Disord.* 16 (2015) 385. Available from: <https://ezp.lib.unimelb.edu.au/login?url=http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=med8&AN=26652166>.
- [50] H. Britt, G.C. Miller, J. Henderson, C. Bayram, C. Harrison, L. Valenti, et al., General Practice Activity in Australia 2015-16, 2016. Available from: <http://www.sydney.edu.au/content/dam/corporate/documents/faculty-of-medicine-and-health/research/research-collaborations,-networks-and-groups/40general-practice-activity-in-australia-2015-16.pdf>.
- [51] M.G. Kahn, T.J. Callahan, J. Barnard, A.E. Bauck, J. Brown, B.N. Davidson, et al., A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data, *eGEMs (Generating Evid Methods to Improv patient outcomes)* 4 (1) (2016 Sep 11) 18. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/27713905>.
- [52] StataCorp LLC, Stata | data analysis and statistical software, Available from: <https://www.stata.com/products/>, 2018.
- [53] M.A. Klebanoff, S.R. Cole, Use of multiple imputation in the epidemiologic literature, *Am. J. Epidemiol.* 168 (4) (2008 Jun 30) 355–357. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18591202>.
- [54] K.J. Lee, G. Roberts, L.W. Doyle, P.J. Anderson, J.B. Carlin, Multiple imputation for missing data in a longitudinal cohort study: a tutorial based on a detailed case study involving imputation of missing outcome data, *Int. J. Soc. Res. Methodol.* 19 (5) (2016 Sep 2) 575–591. Available from: <http://www.tandfonline.com/doi/full/10.1080/13645579.2015.1126486>.
- [55] B.J. Wells, K.M. Chagin, A.S. Nowacki, M.W. Kattan, K.M. Chagin, M.W. Kattan, Strategies for handling missing data in electronic health record derived data, *EGEMS* 1 (3) (2013) 1035. Available from: <http://repository.academyhealth.org/egems>.
- [56] A. Marshall, D.G. Altman, P. Royston, et al., Comparison of techniques for handling missing covariate data within prognostic modelling studies: a simulation study, *BMC Med. Res. Methodol.* 10 (7) (2010), <https://doi.org/10.1186/1471-2288-10-7>.
- [57] P. Peduzzi, J. Concato, E. Kemper, T.R. Holford, A.R. Feinstein, A simulation study of the number of events per variable in logistic regression analysis, *J. Clin. Epidemiol.* 49 (12) (1996 Dec) 1373–1379.
- [58] P. Peduzzi, J. Concato, A.R. Feinstein, T.R. Holford, Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates, *J. Clin. Epidemiol.* 48 (12) (1995) 1503–1510.
- [59] J. Concato, P. Peduzzi, T.R. Holford, A.R. Feinstein, Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy, *J. Clin. Epidemiol.* 48 (12) (1995) 1495–1501. Available from: <https://pubmed.ncbi.nlm.nih.gov/8543963/>.
- [60] J.P. Fine, R.J. Gray, A proportional hazards model for the subdistribution of a competing risk, *J. Am. Stat. Assoc.* 94 (446) (1999) 496–509.
- [61] M. Wolbers, M.T. Koller, J.C.M. Witteman, E.W. Steyerberg, Prognostic models with competing risks methods and application to coronary risk prediction, *Epidemiology* 20 (4) (2009) 555–561.
- [62] J. Hippisley-Cox, C. Coupland, Development and validation of risk prediction algorithms to estimate future risk of common cancers in men and women: prospective cohort study, *BMJ Open* 5 (3) (2015 Mar 17), e007825. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25783428>.
- [63] T.M. Therneau, P.M. Grambsch, Modeling Survival Data: Extending the Cox Model [Internet], Springer New York, New York, NY, 2000 (Statistics for Biology and Health). Available from: <http://link.springer.com/10.1007/978-1-4757-3294-8>.
- [64] D.W. Hosmer, S. Lemeshow, S. May, Applied Survival Analysis, 2nd Edition, Wiley, 2008.
- [65] D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, Applied Logistic Regression, 3rd Edition, Wiley, 2013.
- [66] M.W. Heymans, S. van Buuren, D.L. Knol, W. Van Mechelen, H.C. De Vet, Variable selection under multiple imputation using the bootstrap in a prognostic study, *BMC Med. Res. Methodol.* 7 (33) (2007).
- [67] P. Chondros, S. Davidson, R. Wolfe, G. Gilchrist, C. Dowrick, F. Griffiths, et al., Development of a prognostic model for predicting depression severity in adult primary patients with depressive symptoms using the diamond longitudinal study, *J. Affect. Disord.* 227 (2018 Feb) 854–860. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29689701>.
- [68] D.B. Rubin, Multiple imputation for nonresponse in surveys, in: D.B. Rubin (Ed.), Wiley Series in Probability and Statistics, Wiley InterScience (Online service), Hoboken, NJ, USA, 1987, p. 258, <https://doi.org/10.1002/9780470316696>. Available from:.
- [69] M.W. Kattan, J. Marasco, What is a real nomogram? *Semin. Oncol.* 37 (1) (2010 Feb) 23–26. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0093775409002280>.
- [70] S. Makridakis, E. Spiliotis, V. Assimakopoulos, Statistical and Machine Learning forecasting methods: concerns and ways forward, Hernandez Montoya AR, *PLoS One* 13 (3) (2018 Mar 27), e0194889. Available from: <https://dx.plos.org/10.1371/journal.pone.0194889>.
- [71] Research I of M (US) C on S for SR of CE, J. Eden, L. Levit, A. Berg, S. Morton, Finding what Works in Health Care: Standards for Systematic Reviews, Finding What Works in Health Care [Internet], National Academies Press (US), 2011. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24983062>.
- [72] D. Moher, A. Liberati, J. Tetzlaff, D.G. Altman, D. Altman, G. Antes, et al., Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement, *PLoS Med.* 6 (2009).
- [73] CASP UK, CASP Checklists - CASP - Critical Appraisal Skills Programme, 2020. Available from: <https://casp-uk.net/casp-tools-checklists/>.
- [74] Sax Institute, Sure - Sax Institute - Sax Institute, Available from: <https://www.saxinstitute.org.au/our-work/sure/>.
- [75] Health and Biomedical Informatics Centre TU of M, The university of Melbourne health and biomedical Informatics Centre. What is GRHANITE™?, Available from: <https://www.grhanite.com/technologies/>, 2016.
- [76] L. Nguyen, M. Stooze, D. Boyle, D. Callander, H. McManus, J. Asselin, R. Guy, B. Donovan, M. Hellard, C. El-Hayek, et al., Privacy-preserving record linkage of deidentified records within a public health surveillance system: evaluation study (Preprint), *J. Med. Internet Res.* 22 (6) (2020) e16757.
- [77] Australian Medicines Handbook Pty Ltd, Australian Medicines Handbook (Online), 2019.
- [78] NPS MedicineWise, General practice insights report june 2017-june 2018, Sydney (2019). Available from: <https://www.nps.org.au/assets/NPS/pdf/General-Practice-Insights-Report-2017-18.pdf>.
- [79] Statistics AB of. Australian statistical Geography standard (ASGS) [Internet]. Available from: ++++[https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+\(ASGS\)](https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS)).