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 health record
Electronic medical 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...

* 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.jocarto.2020.100126
Received 16 July 2020; Accepted 17 November 2020
2665-9131/© 2020 The Authors. Published by Elsevier Ltd on behalf of Osteoarthritis Research Society International (OARSI). This is an open access article under the
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].

### Table 1

| 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/m²                       |
| 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: |
| 1 – Diagnosis of anxiety, depression, post-traumatic stress disorder (PTSD), obsessive compulsive disorder (OCD), anorexia, bulimia, bipolar disorder, dissociative disorders or schizophrenia |             |                                      | - Charlson Comorbidity Index (CCI) [28] |
| 1 = Did not undergo TKR                            |             |                                      |                             |
| 1 = Underwent contralateral/previous TKR           | AOANJRR     | Binary                               | 0 = Did not undergo contralateral/previous TKR |
| 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                               | 0 = Did not undergo contralateral/previous TKR |
| Any past knee surgery on either knee               | NPS EMRs    | Binary                               | 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: |
| 1 = Inner regional                                 |             |                                      | 0 = Major cities            |
| 2 = Outer regional                                 |             |                                      | 1 = Inner regional          |
| 3 = Remote                                        |             |                                      | 2 = Outer regional          |
| 4 = Very remote                                    |             |                                      | 3 = Remote                  |
| **Model two**                                      |             |                                      |                             |
| Non-response to TKR (outcome)                       | SMART       | Binary                               | 1 = Non-responder to TKR (as per OMERACT OARSI criteria) |
| BMI                                                | SMART       | Continuous                           | 0 = Responder to TKR         |
| Multimorbidity                                     | SMART       | Continuous                           | Kg/m²                       |
| Mental health condition                            | SMART       | Binary                               | 1 = Diagnosis of anxiety, depression, bipolar disorder, schizophrenia, post-traumatic stress disorder, borderline personality disorder or obsessive-compulsive disorder |
| Use of opioids prior to surgery                     | MBS/PBS     | Categorical                          | 1 = Use of non-opioid osteoarthritis medication only |
| Use of fibromyalgia                                | SMART       | Binary                               | 0 = Use of opioid osteoarthritis medication only |

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...
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 MedicinesInsight 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 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 MedicinesInsight 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
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].
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.
models. 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 included for each predictor as the proportion of influential outliers.

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 × 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 models. 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 2

| 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 |
| | Logistic regression |
| Model estimation | Maximum log-likelihood and shrinkage (bootstrapping) |
| Model performance | Discrimination- Harrell's overall C statistic |
| | Discrimination- Concordance C statistic |
| Model validation | Calibration- predicted and observed 5-year risk |
| | Calibrations- observed probabilities of non-response |
| Model presentation | Regression equation, nomogram and web-based tool (future) |

The proportional hazards assumption will be checked during model one development.

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 correction 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.

### Table 3

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

Abbreviations: CI = confidence interval.

4.7.7. 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.
Inconsistent 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) → | MeSH (knee replacement/arthroplasty) | MeSH (osteoarthritis/knee osteoarthritis) | MeSH (models/statistical) model* | MeSH (nomograms) nomogram* |
|---------|-------------------------------------|-----------------------------------------------|---------------------------------|--------------------------|
| (OR)    | knee replacement*                   | knee osteoarthritis                           | models*                         | nomograms*               |
|         | knee arthroplast*                   |                                                |                                 |                          |

Limit: English language and last 10 years.
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. | 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. |
| 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. | 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. |
| 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. | 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. |
| 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. | 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. |
| 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 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. |
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

### Table A 1.3
Summary of prediction models for TKR

| Study                  | Outcome                              | Setting                                                                                     | Factors                                                                 |
|------------------------|--------------------------------------|---------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Lewis et al. (N = 1462) | 10 yr risk of TKR in Australian women ≥70 years with OA | Data linkage study using hospital discharge data throughout Western Australia and data from a randomised controlled trial | Sociodemographic factors: Age (years) 1.00 (0.93–1.07) 1.13 ** (1.01–1.07) |
| Zeni et al. (N = 120)   | TKR 2yrs post referral to orthopaedic surgeon in American patients ≥46 yrs with OA | University of Delaware Physical Therapy Clinic                                              | Clinical factors: BMI (kg/m²) 1.07 (1.03–1.11)                        |
| Hawker et al. (N = 2128)| 5 yr risk of TJR in Canadian patients aged ≥55 years with OA of knee or hip | Population study conducted in two regions in Ontario, Canada                              | Knee osteoarthritis history: Prevalent knee replacement/past knee surgery 2.08 (1.01–4.28) |
| Riddle et al. (N = 4670)| TKR at 3 yr follow-up in American patients aged 45–79 years with OA or at risk of OA | Multicentre (1 hospital and 3 university sites) community-based study conducted in Maryland, Rhode Island, Ohio and Pennsylvania | Knee pain and symptoms: Knee pain (infrequent, frequency, daily vs no) 2.35 (1.86–2.97) |

| Study                  | Final model OR (95% CI) | Final model OR (95% CI) | Final model HR (95% CI) | Final model RR (95% CI) | Final model OR (95% CI) | Final model OR (95% CI) | Final model HR (95% CI) | Final model RR (95% CI) |
|------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| Lewis et al. (N = 1462)| 1.00 (0.93–1.07)        | 1.13 ** (1.01–1.07)     | 1.04 (1.01–1.07)        | 2.04 (3.33–3.13)        | 1.98 (1.33–2.95)        | 1.22 (1.13–1.31)        | 1.64 (0.87–3.12)        |
| Zeni et al. (N = 120)  | 1.07 (1.03–1.11)        | 2.08 (1.01–4.28)        | 2.04 (1.33–3.13)        | 1.81 (1.12–2.92)        | 2.35 (1.86–2.97)        | 1.58 (1.04–2.39)        | 1.58 (1.04–2.39)        |
| Hawker et al. (N = 2128)| 1.06 (1.02–1.11)        | 0.71 (0.43–1.18)        | 0.71 (0.43–1.18)        | 0.71 (0.43–1.18)        | 1.07 (1.03–1.11)        | 1.06 (1.02–1.11)        | 1.06 (1.02–1.11)        |
| Riddle et al. (N = 4670)| 1.06 (1.02–1.11)        | 0.71 (0.43–1.18)        | 0.71 (0.43–1.18)        | 0.71 (0.43–1.18)        | 1.07 (1.03–1.11)        | 1.06 (1.02–1.11)        | 1.06 (1.02–1.11)        |

** 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 | Outcome | Setting | Factors | Sociodemographic factors | Clinical factors | Measures of mental health | Measures of function and pain |
|-------|---------|---------|---------|--------------------------|------------------|----------------------------|-------------------------------|
| Dowsey et al. (N = 615) | Non-response to TKR (defined using OARSI responder criteria) 12 months post-surgery in Australian patients | Data from an arthroplasty registry at a Melbourne hospital | Non-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 | Age 2.66 (2.61–2.71) | Presence of low back pain BMI (≥40 kg/m²) v.s. ≥ 40 kg/m² | 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 | WOMAC summary score 0.81 (0.68–0.97) 1.32 (1.23–1.42) 5 WOMAC questions 0.58 (0.50–0.67) |
| Hawker et al. (N = 202) | Response to TJR (defined as change in WOMAC summary score) 12 months post-surgery in Canadian patients aged 55 years and older with at least moderately severe knee/hip arthritis | Population study conducted in two regions in Ontario, Canada | Population study conducted in two regions in Ontario, Canada | Final model OR (95% CI) | Presence of low back pain BMI (≥40 kg/m²) v.s. ≥ 40 kg/m² | 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 | WOMAC summary score 0.81 (0.68–0.97) 1.32 (1.23–1.42) 5 WOMAC questions 0.58 (0.50–0.67) |
| Razak et al. (N = 3062) | Response to TKR (defined as change in WOMAC score) 12 months post-surgery in UK patients | Multicentre (34 centres) randomised controlled trial in the United Kingdom | Multicentre (34 centres) randomised controlled trial in the United Kingdom | Final model OR (95% CI) | Presence of low back pain BMI (≥40 kg/m²) v.s. ≥ 40 kg/m² | 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 | WOMAC summary score 0.81 (0.68–0.97) 1.32 (1.23–1.42) 5 WOMAC questions 0.58 (0.50–0.67) |
| Langu et al. (N = 141) | 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 | Data from a tertiary institution in Singapore | Data from a tertiary institution in Singapore | Final model OR (95% CI) | Presence of low back pain BMI (≥40 kg/m²) v.s. ≥ 40 kg/m² | 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 | WOMAC summary score 0.81 (0.68–0.97) 1.32 (1.23–1.42) 5 WOMAC questions 0.58 (0.50–0.67) |
| Clement et al. (N = 966) | Poor response to TKR (WOMAC score in last quintile) 6 months post-surgery in Canadian patients aged ≥40 years | Multicentre (3 hospitals) study in Quebec City, Canada | Multicentre (3 hospitals) study in Quebec City, Canada | Final model Mean difference in change in OKS score (95% CI) | Presence of low back pain BMI (≥40 kg/m²) v.s. ≥ 40 kg/m² | 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 | WOMAC summary score 0.81 (0.68–0.97) 1.32 (1.23–1.42) 5 WOMAC questions 0.58 (0.50–0.67) |

Abbreviations: Standard deviation (SD), Total Knee Replacement (TKR), Osteoarthritis Research Society International (OARSI), Osteoarthritis (OA), Kellgren and Lawrence (K-L), 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), Knee Score 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 (BMI ≥40 kg/m² compared with <40 kg/m²) 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 (63) | 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) | |
Table B 1.1 (continued)

| Condition                        | Yes (EMR) | No (SMART) | 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/SMART 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/m$^2$. 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

| 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

| ATC code | Drug name                                      |
|----------|-----------------------------------------------|
| H02AB01  | betamethasone                                 |
| H02AB02  | dexamethasone                                 |
| H02AB04  | methylprednisolone                            |
| H02AB06  | prednisolone                                  |
| H02AB07  | prednisone                                    |
| H02AB08  | triamcinolone                                 |
| H02AB09  | hydrocortisone                                |
| H02AB10  | cortisone                                     |
| H02AB15  | meprednisone                                  |
| H02B0X1  | methylprednisolone, combinations              |
| M01AB01  | indomethacin                                  |
| M01AB02  | sulindac                                      |
| M01AB05  | diclofenac                                    |
| M01AB15  | ketorolac                                     |
| M01AB51  | indomethacin, combinations                    |
| M01AB55  | diclofenac, combinations                      |
| M01AC01  | piroxicam                                     |
| M01AC06  | meloxicam                                     |
| M01AC56  | meloxicam, combinations                       |
| M01AE01  | ibuprofen                                     |
| M01AE02  | Naproxen                                      |
| M01AE03  | ketoprofen                                    |
| M01AE14  | dexibuprofen                                  |
| M01AE17  | dexketoprofen                                  |
| M01AES1  | ibuprofen, combinations                       |
| M01AES2  | naproxen and esomeprazole                    |
| M01AES3  | ketoprofen, combinations                      |
| M01AES6  | 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)
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.

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      | capsicain and similar agents       |
| M02AB01    | Capsicain                          |
| M02AB02    | zucapsicain                        |
| M09AX01    | hyaluronic acid                    |
| N01BX04    | Capsicain                          |
| N02AA01    | Morphine                           |
| N02AA03    | hydromorphone                      |
| N02AA05    | oxycodone                          |
| N02AA08    | dihydrocodeine                     |
| N02AA51    | morphine, combinations             |
| N02AA53    | hydromorphone and naloxone         |
| N02AA55    | oxycodone and naloxone             |
| N02AA56    | oxycodone and naltrazone           |
| 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                         |
Coding methodology for OA medications

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

| Chronic conditions most frequently managed in primary care from BEACH study |
|-----------------------------|-----------------------------|
| Charlson Comorbidity Index (CCI) |                  |
| 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) | 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   | Peripheral vascular disease                       |
|                               | Diabetes vascular disease                         |
|                               | Arteritis-diabetes mellitus                       |
|                               | Buerger’s disease                                 |
|                               | Obliterative vascular disease                     |
|                               | Diabetic endarteritis                             |
|                               | Thrombangitis obliterans                          |
|                               | Arteriosclerosis obliterans                       |
|                               | Diabetes with vascular changes                    |
|                               | Occlusive vascular disease                        |
|                               | Peripheral arterial disease                       |
| CVA/TIA                       | Cerebral haemorrhage                              |
|                               | Cerebral infarction                               |
|                               | Cerebrovascular accident                          |
|                               | 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)
| Condition                        | Free text terms                                                                 |
|---------------------------------|---------------------------------------------------------------------------------|
| **Peptic ulcer disease**        | Polymyositis                                                                     |
|                                 | Dermatomyositis                                                                  |
|                                 | Peptic ulcer                                                                     |
|                                 | Ulcerative reflux disease                                                        |
|                                 | Gastric ulcer                                                                    |
|                                 | Stomach ulcer                                                                    |
| Liver disease                   | Gerd with ulceration                                                             |
|                                 | 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                                                              |
| Liver disease                   | Drug induced hepatitis                                                            |
|                                 | Hemiplegia                                                                       |
|                                 | Leukaemia                                                                        |
|                                 | Lymphoma                                                                         |
| Malignant tumour                | Cancer                                                                           |
|                                 | Malig                                                                            |
|                                 | Carcinoma                                                                        |
|                                 | 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/p                                                                              |
|                                 | Hypertensive                                                                     |
|                                 | Hm                                                                               |
|                                 | Bp high                                                                          |
|                                 | Labile bp                                                                        |
|                                 | Bp labile                                                                        |
|                                 | Labile blood pressure                                                            |
|                                 | Blood pressure labile                                                            |
|                                 | Hbp                                                                              |
|                                 | Elevated blood pressure                                                          |
|                                 | Raised blood pressure                                                            |
|                                 | Antihypertensive agent                                                           |
|                                 | Hbpm                                                                             |
| **Diabetes mellitus**           | T2dm                                                                             |
|                                 | Diab                                                                             |
|                                 | Niddm                                                                            |
|                                 | Mellitus                                                                         |
|                                 | Ketoac                                                                           |
|                                 | Dka                                                                              |
|                                 | Osmolar                                                                          |
|                                 | Insulin                                                                          |
| **Lipid disorder**              | Hyperlip                                                                          |
|                                 | Hypercholesterol                                                                  |
|                                 | Hypertrig                                                                        |
|                                 | High chol                                                                        |
|                                 | Hyperlipoprotein                                                                  |
|                                 | High lipid                                                                       |
|                                 | Dyslip                                                                           |
| **Dyspepsia/oesophageal disease**| Gor                                                                              |
|                                 | Hb                                                                               |
|                                 | Heartburn                                                                        |
|                                 | Dyspepsia                                                                        |
|                                 | Reflux                                                                           |
|                                 | Barrett                                                                          |
|                                 | Indigest                                                                         |
|                                 | Petic ulcer                                                                       |
|                                 | Hyperacidity                                                                      |
|                                 | Sliding hernia                                                                    |

(continued on next column)
| Condition | Free text terms |
|-----------|-----------------|
| Regurgitation | Oesophagitis |
| Hiatus hernia | Epigastric pain |
| Belching | Helicobacter pylori |
| Asthma | Asthma |
| Samter | Wheezy bronchitis |
| Asthmaticus | Asthmoid |
| Atrial fibrillation | Af |
| A/f | Atrial fibr |
| A.F. | A FIB |
| Malignant neoplasms of the skin | Scc |
| BCC | Squamous cell |
| Basal cell | Melanoma |
| Osteoporosis | Osteoporosis |
| Osteoporotic | Hypothyroid |
| Hypothyroidism/myxoedema | Myxoedema |
| Hashimotos | Thyroiditis |
| Ischaemic heart disease (ihd) | 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 |
| Pts | 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 |
| Copd | Emphysema |
| Emphysematous | Chronic airflow/s limitation |
| Chronic bronchitis | Chronic obstructive airway |
| Chronic obstructive pulmonary | Bronchiitis- chronic |
| Chronic airways disease | Chronic airflow limitation |
| Chronic airflow/s infection | Cal |
| Chronic obstructive pulmonary disease | Bronchitis chronic |
| Gout | Gout |
| Migraine | Migraines |
| Migratory | Congestive heart failure |
| Heart failure | (continued on next column) |
Table D 1.4 (continued)

| Condition                  | Free text terms                      |
|----------------------------|--------------------------------------|
| Cardiac failure            | Cardiac failure                      |
|                            | Ccf                                  |
|                            | Chf                                  |
|                            | Cor pulmonae                         |
|                            | Lhf                                  |
|                            | Rhf                                  |
|                            | Lvf                                  |
|                            | Rvf                                  |
|                            | Cardiac dysfunction                  |
|                            | Cardiomyopathy                       |
|                            | Ventricular failure                  |
|                            | Ventricular diastolic dysfunction    |
| Chronic pain               | Headache                             |
|                            | Plantar fascitis                     |
|                            | Carpal tunnel                        |
|                            | Rheumatoid arthritis                 |
|                            | Dysarthrosis                         |
|                            | Shingles                             |
|                            | Neuralgia                            |
|                            | Compressed nerve                     |
|                            | Spondylolisthesis                    |
|                            | Back pain                            |
|                            | Back problem                         |
|                            | Neck pain                            |
|                            | Neck problem                         |
|                            | Thoracic pain                        |
|                            | Thoracic problem                     |
|                            | Sciatica                             |
|                            | Bulging disc                         |
|                            | Dementia                             |
|                            | Demented state                       |
|                            | Dementing illness                    |
|                            | Dementio vascular                    |
|                            | Alzheimers                           |
| CHRONIC KIDNEY DISEASE     | 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, ANTIANXIETY 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                | BI-POLAR                                      |
| Schizophrenia           | Schizophrenia                                 |
| 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            | Menisceotomy                        |
| 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 | Arthroscopy                        |
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/ostearthritiss/contents/what-is-osteoarthritis.
[2] I.N. Ackerman, C. Pratt, A. Gorelik, D. Liew, Projected burden of osteoarthritis and hip and knee osteoarthritis, Australian Institute of Health and Welfare, 2017 [Internet]. AIHW website. Available from: https://www.aihw.gov.au/reports/chronic-musculoskeletal-conditions/ostearthritiss/what-is-osteoarthritis.
[3] Australian Institute of Health and Welfare, Osteoarthritis and Cartilage Open 2 (2020) 100126.
Which patients are most likely to benefit from total joint arthroplasty? Arthritis Rheum. 65 (5) (2013) 1243–1252. Available from: https://pubmed.ncbi.nlm.nih.gov/23995200.

"K.G.M. Moons, P. Royston, Y. Vergouwe, D.E. Grobbee, D.G. Altman, Prognosis and causal inference from the northern OA study, Osteoarthr Cartil 25 (Suppl: 1) (2017) 742S–750S. Available from: https://pubmed.ncbi.nlm.nih.gov/28757187.

"H.R.B.A. Razak, C.S. Tan, Y.J.D. Chen, H.N. Pang, K.J. Darren Tay, P.L. Chin, et al., Socioeconomic status affects the Oxford knee score and short-form 12 score in Singapore. Ann. Rheum. Dis. 70 (12) (2011) 2198–2204. Available from: https://pubmed.ncbi.nlm.nih.gov/21819833.

"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://pubmed.ncbi.nlm.nih.gov/25950705.

"C.M. Hochberg, N.I. Wood, et al., Which patients are most likely to benefit from total joint arthroplasty? Arthritis Rheum. 65 (5) (2013) 1243–1252. Available from: https://pubmed.ncbi.nlm.nih.gov/23995200.

"S. Thuraisingam et al. Osteoarthritis and Cartilage Open 2 (2020) 100126
