

BMJ Open Development of prediction models for complications after primary total hip and knee arthroplasty: a single-centre retrospective cohort study in the Netherlands

Lieke Sweerts ^{1,2}, Thomas J Hoogeboom,² Thierry van Wessel,¹ Philip J van der Wees,^{2,3} Sebastiaan A W van de Groes¹

To cite: Sweerts L, Hoogeboom TJ, van Wessel T, *et al*. Development of prediction models for complications after primary total hip and knee arthroplasty: a single-centre retrospective cohort study in the Netherlands. *BMJ Open* 2022;**12**:e062065. doi:10.1136/bmjopen-2022-062065

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-062065>).

Received 16 February 2022
Accepted 23 July 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Lieke Sweerts;
lieke.sweerts@radboudumc.nl

ABSTRACT

Objective The aim of this study was to develop prediction models for patients with total hip arthroplasty (THA) and total knee arthroplasty (TKA) to predict the risk for surgical complications based on personal factors, comorbidities and medication use.

Design Retrospective cohort study.

Setting Tertiary care in outpatient clinic of university medical centre.

Participants 3776 patients with a primary THA or TKA between 2004 and 2018.

Primary and secondary outcome

measures Multivariable logistic regression models were developed for primary outcome surgical site infection (SSI), and secondary outcomes venous thromboembolism (VTE), postoperative bleeding (POB), luxation, delirium and nerve damage (NER).

Results For SSI, age, smoking status, body mass index, presence of immunological disorder, diabetes mellitus, liver disease and use of non-steroidal anti-inflammatory drugs were included. An area under the receiver operating characteristic curve (AUC) of 71.9% (95% CI=69.4% to 74.4%) was found. For this model, liver disease showed to be the strongest predictor with an OR of 10.7 (95% CI=2.4 to 46.6). The models for POB and NER showed AUCs of 73.0% (95% CI=70.7% to 75.4%) and 76.6% (95% CI=73.2% to 80.0%), respectively. For delirium an AUC of 85.9% (95% CI=83.8% to 87.9%) was found, and for the predictive algorithms for luxation and VTE we found least favourable results (AUC=58.4% (95% CI=55.0% to 61.8%) and AUC=66.3% (95% CI=62.7% to 69.9%)).

Conclusions Discriminative ability was reasonable for SSI and predicted probabilities ranged from 0.01% to 51.0%. We expect this to enhance shared decision-making in considering THA or TKA since current counselling is predicated on population-based probability of risk, rather than using personalised prediction. We consider our models for SSI, delirium and NER appropriate for clinical use when taking underestimation and overestimation of predicted risk into account. For VTE and POB, caution concerning overestimation exceeding a predicted probability of 0.08 for VTE and 0.05 for POB should be taken into account. Furthermore, future studies should

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This study included multivariable logistic regression models to predict postoperative complications after primary total hip and knee arthroplasty based on personal factors, comorbidities and medication use.
- ⇒ The present study was conducted and reported according to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis guidelines.
- ⇒ Purposive selection of predictors by clinical reasoning and literature search.
- ⇒ Limitations include only internal validation of the prediction models by bootstrapping.
- ⇒ Used data were not primarily registered for research purposes, and therefore, their detail and accuracy could be less than optimal.

evaluate clinical impact and whether the models are feasible in an external population.

INTRODUCTION

Joint replacement is a recommended intervention for people with end-stage hip or knee osteoarthritis.¹ Whether surgery is the best solution depends on many individual factors such as severity of the disease, level of experienced pain and discomfort, medication use, personal circumstances, comorbid diseases and intended type of surgery.²⁻⁴ Because the decision to have surgery or not is complex, a shared decision-making (SDM) process is warranted. This process allows patients and clinicians to discuss treatment options consistent with the patient's values and preferences.⁵

Information on most likely prognosis is central in this dialogue as the clinician provides guidance and information about expected outcomes, including the risk on surgical complications, when facing the

decision to pursue or forgo surgery. However, providing personalised information about the risk on surgical complications, based on personal characteristics of the patient, is challenging. Available evidence often consists of average outcomes and current guidelines on prediction of outcome still recommend counselling predicated on population-based probability of risk, rather than using personalised prediction.⁶ This is remarkable, as discussing potential personal risks is an important aspect of SDM.^{7,8}

To overcome this problem, models that can predict postoperative complications are frequently developed and applied. Several universal surgical prediction models have already been developed based on a big national database.⁹ However, before applying these models to orthopaedic surgical procedures, performance and accuracy on the specific surgical field needs to be determined. For total joint arthroplasty, this is performed by Trickey *et al.*¹⁰ As shown by Trickey *et al.*, and others, patients at risk of not benefitting from total hip arthroplasty or total knee arthroplasty (THA or TKA) can be identified using prediction models based on preoperative data like demographic factors, and pain scores, and physical functioning measured with patient reported outcome measures.¹⁰⁻¹³ Another study developed a preoperative prediction model to predict residual reports on pain, functional outcome and treatment success for individual patients after TKA.¹⁴ Also useful electronic risk calculators predicting complications and mortality for patients and clinicians are available for specific populations.¹⁵⁻¹⁷ In one study, data of patients registered in the Medicare database, the federal health insurance programme for individuals aged ≥ 65 years, are used for development of a risk calculator. However, the exact patient characteristics of the study population are not reported and the effect of the predictors remain unclear.¹⁶ Harris *et al.* developed prediction models with machine learning techniques models to determine demographic and clinical predictors for prediction of postoperative complications and mortality. The authors were able to identify predictor variables for their three most accurate models predicting a postoperative renal complication, cardiac complication and death. However, used predictor variables in the models can only be found for their three most accurate outcomes.¹⁷ Further research is warranted to identify relevant predictors for different postoperative outcomes. In another study, regression models are based on the results of univariate analyses on a broad range of data as demographics, comorbidities and laboratory, or test values of a mainly male veteran population, and the authors reported suboptimal performance scores for prediction of most outcomes.¹⁵ Generalisability of prediction models based on specific patient populations may be limited, and further evaluation of potential risk factors is needed to validate prediction models for complications after primary total hip and knee replacement.

As it is known from literature that personal factors including demographic characteristics and comorbidities have an impact on surgical complications,³ these assumed

caused relationships might therefore serve as basis for a risk prediction model. Therefore, the aim of this study is to develop a prediction model for clinicians and patients with hip or knee osteoarthritis considering surgery, by predicting risk for surgical complications based on personal factors, comorbidities and medication use.

METHODS

Study design and setting

For this retrospective cohort study, we established a cohort of patients who underwent primary THA or TKA between 2004 and 2018 at the Orthopaedic Department of Radboud university medical center Nijmegen, the Netherlands. Data sets were merged into one centralised database based on patient number, birthdate and date of surgery.

This study was performed and reported in line with transparent reporting of a multivariable prediction model for individual prognosis or diagnosis guidelines.¹⁸

Data collection

Data used for this study were extracted from (electronic) medical records of Radboudumc, Dutch Arthroplasty Register (LROI) and Radboudumc registry of complications. We primarily extracted comorbidities and medication use from medical records. These data were extracted based on coding and were obtained by three researchers (LS, TvW and AT) by use of a standardised operating procedure, and stored in a centralised platform (Castor Electronic Data Capture).¹⁹ Data about patient characteristics like age, sex, body mass index (BMI), smoking status, American Society of Anesthesiologists (ASA) classification and diagnosis for surgery were extracted from LROI. Furthermore, date of surgery, type of surgery (primary or revision), surgery side and type of implant were extracted.²⁰ From the register of complications we extracted all surgeries and complications which occurred within 1 year after THA or TKA.²¹ In this registry, surgery-related orthopaedic complications were registered as well as other medical complications.²² All complications were registered by location code combined with a code for the nature of the complication.²¹ Some registrations were unclear and could refer to one of predefined complications and were therefore checked in medical records by LS. For all included location and nature of complication codes per surgical complication, see online supplemental eTable 1.

Inclusion and exclusion criteria

Patients were eligible for inclusion in the cohort if the surgery concerned primary THA or TKA. We defined primary THA or TKA as the first time a total prosthesis is placed. Revision arthroplasty was defined as any change (replacement, removal or addition) of one or several components of the joint prosthesis.²⁰ We expected revision arthroplasty to influence risk for complications

negatively, therefore revision arthroplasty was excluded for this study.

Outcome (dependent variables)

Prediction models were developed over the pooled THA and TKA data for six predefined surgical complications. Primary outcome was surgical site infection (SSI), and secondary outcomes included venous thromboembolism (VTE), postoperative bleeding (POB), luxation, delirium and nerve damage (NER). All prediction models were developed based on primary THA and TKA data, except for the models for luxation and NER which were developed based on primary THA data. These surgical complications are uncommon in TKA.

Predictors (independent variables)

In total 16 predictor candidates were selected based on evidence from previous reports and clinical reasoning in relation to the outcomes. These included patient characteristics, comorbidities and medication use (as specified in online supplemental eTables 2 and 3). Note that we made a purposive selection from the 16 predictors candidates to serve as predictors for the different surgical complications.

Comorbidities extracted from medical records were categorised according to the English National Health Service (NHS). The NHS considered these categories relevant comorbid categories in terms of outcome prediction.³ Medication use was reduced to the active substance of the drug and was categorised to drug groups according to the Dutch pharmacotherapeutic compass.²³

Sample size

It is recommended that at least five events are collected for each predictor that is evaluated in multivariable regression analysis.^{24 25} An event was defined as the least frequent outcome status, which in our case was the presence of surgical complication. In the Netherlands, the estimated risk of a complication like SSI is 3%²⁶; therefore, in order to develop a model with six predictors, at least 30 events were required, and so a sample size of at least 1000 patients was required.

Missing data

Data were checked for completeness by investigating patterns of missingness to assess presence of a non-random element. Incomplete data were double-checked. Missing data were imputed using multiple imputation, as the omission of patients who have one or more predictor variables missing from analysis can cause considerable loss of precision and might bias the results.^{27 28} The number of imputations was set to 10. The imputation was checked for accuracy by visual inspection and frequencies.

Statistical analysis methods

Model development

Evidence from literature, clinical reasoning and eyeballing guided selection of predictors to be included in the models. Eyeballing was done by evaluation of

potential higher frequencies of predictors in relation to the outcome.²⁹ All selected predictors were entered into a multivariable logistic regression model, using the occurrence of a surgical complication as outcome variable. The prediction model was pooled over the imputed data sets.³⁰

Internal validation

To reduce risk of overfitting, we internally validated the model using bootstrapping. In this step, B-bootstrap samples of B=1000 were drawn with replacement from original data, which reflects drawing samples from underlying population. Due to the drawing with replacement, a bootstrapped data set allows for containing the same original cases. Other validation methods resample without replacement and thereby such validation data sets are produced through a prespecified number of surrogate data sets, and each of the original cases will be left out exactly once, which results in a smaller data set. Since our data set is not very large, we decided to use bootstrapping as internal validation method. Bootstrapping was performed to estimate the performance in future patients, and to adjust the model by the calculated shrinkage factor so that future predictions will be less extreme.²⁴

Performance of the model

We quantified measures of performance, discrimination and calibration. Overall model performance is the distance between predicted and actual outcome.²⁸ To quantify overall model performance, we assessed Brier, Brier_{scaled} and Nagelkerke's R². For Brier, squared differences between actual outcome and predictions were calculated. Brier can range from 0 for a perfect model to 0.25 for a non-informative model with 50% incidence of the outcome. Brier_{scaled} is scaled by its maximum under a non-informative model and range between 0% and 100%. Nagelkerke's R² is a measure of explained variation.³¹ The ability of the model to discriminate between those with and without the outcome was quantified as the area under the receiver operating characteristic curve (AUC). This can range from 50% (no discriminative capacity) to 100% (perfect discriminative capacity). The discriminative capacity was interpreted as reasonable when AUC was >0.70 and good when AUC was >0.80.³² Calibration of the model is the agreement between predicted probabilities (probability of an event calculated with the model) and observed frequencies of outcome (accuracy) and was assessed by visually inspecting the calibration plot.²⁸ Furthermore, we computed Hosmer and Lemeshow (H-L) goodness-of-fit as a quantitative measure of calibration. A high H-L statistic is related to a low p value, and indicates a poor fit.²⁴

All statistical analyses were performed using R V.3.5.3. Packages *vim*, *mice*, *rms*, *pROC* and *generalhoslem* were used.

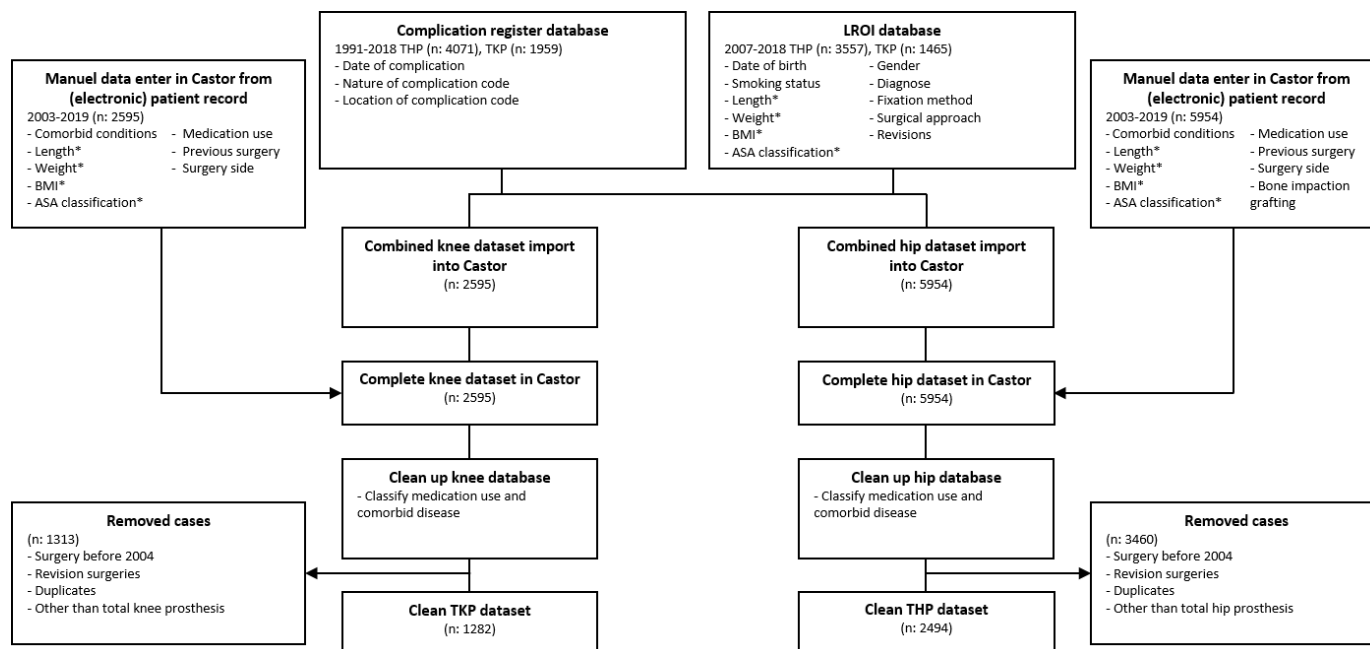


Figure 1 Flow chart for inclusion and exclusion of patients. Variables indicated with an asterisk* were primarily extracted from the LROI database. When these data were missing, the data were extracted from the (electronic) medical record. Castor Electronic Data Capture is indicated by Castor. ASA, American Society of Anesthesiologists; BMI, body mass index; LROI, Dutch Arthroplasty Register; THP, total hip replacement; TKP, total knee replacement.

Patient and public involvement

Patients were involved in the design of the study which included consultation during grant writing and advice in setting up the study design. Furthermore, patients were involved in the process of incorporating the prediction models in a patient decision aid. Focus groups were held and patients and clinicians together were asked for their opinion regarding incorporation of the models in the preoperative process.

RESULTS

Participants

In total 3776 patients with primary THA or TKA were identified as eligible for the present study. Of these patients, 2494 patients underwent THA and 1282 patients underwent TKA. See figure 1 for participant flow. Baseline characteristics of the final cohort are presented in table 1.

Model development

The number of missing values per predictor are shown in table 1. For the majority of potential predictors, there was only a small quantity of missing data; however, smoking status was missing in 24.7%. After imputation, all patients were available for multivariable modelling. There were no missing values in surgical complications.

Model specification

According to our selection of predictor candidates per outcome (depicted in online supplemental eTable 4), we entered all selected predictors in the model. For SSI, these predictors were: age, smoking status, BMI, presence of an immunological disorder, diabetes mellitus, liver

disease and use of non-steroidal anti-inflammatory drugs. We found a significant influence of age, immunological disorder, diabetes mellitus and liver disease of which the presence of liver disease showed to be the strongest predictor with an OR of 10.7 (95% CI=2.4 to 46.6). The bootstrap yielded a shrinkage factor of 0.984, which was used to adjust the regression coefficients. Table 2 shows the adjusted prediction models and ORs that estimates the risk for SSI and secondary outcomes. For original prediction models and adjusted coefficients, see online supplemental eTable 5.

Model performance

Brier, Brier_{scaled} and Nagelkerke's R^2 , to assess overall performance of the model for SSI, were 0.010, 0.026 and 0.081, respectively.

The discriminative performance of the model for SSI is shown in figure 2. The AUC was 71.9% (95% CI=69.4% to 74.4%), which indicates reasonable discriminative ability. Predicted probabilities ranged between 0.01% and 51.0%, with a mean of 1.0% (SD=1.5%). Calibration was poor, indicated by significant H-L statistic ($p<0.001$). The corresponding calibration plot that represents the accuracy of the model is shown in figure 3. The calibration plot showed quite accurate prediction, especially when the risk is low. The model underestimates the risk with a predicted probability >0.10 .

The performance, discrimination and calibration of SSI and secondary outcomes are presented in table 3. The predictive algorithms for POB and NER showed reasonable discriminative values (AUC=73.0 and 76.6) and explained fraction of variance by a Nagelkerke's R^2

Table 1 Patient characteristics

Patient characteristics	Missing values	Total population (n=3776)	Total hip replacement (n=2494)	Total knee replacement (n=1282)
Age, mean (SD), years	0.10%	60.2 (15.8)	57.7 (17.0)	65.1 (11.7)
Gender: female, no. (%)	0.10%	2298 (60.9)	1468 (58.9)	829 (64.7)
BMI, mean (SD), kg/m ²	2.60%	27.5 (5.2)	26.6 (4.7)	29.3 (5.6)
Smoking: yes, no. (%)	24.70%	498 (13.2)	341 (13.7)	157 (12.2)
ASA classification, no. (%)	0.40%			
I		839 (22.2)	669 (26.8)	170 (13.3)
II		2091 (55.4)	1314 (52.7)	777 (60.6)
III		829 (22.0)	500 (20.0)	329 (25.7)
Diagnosis hip, no. (%)	0.40%			
Arthrosis			1599 (64.1)	
Rheumatoid arthritis			68 (2.7)	
Dysplasia			241 (9.7)	
Osteonecrosis			228 (9.1)	
Other			349 (14.0)	
Diagnosis knee, no. (%)	0.90%			
Arthrosis				1037 (80.9)
Rheumatoid arthritis				123 (9.6)
Other				111 (8.7)
Side affected: right, no. (%)	0.30%	1915 (50.9)	1257 (50.4)	658 (51.3)
Surgical complications, no. (%)	0%			
Surgical site infection		38 (1.0)	25 (1.0)	13 (1.0)
Venous thromboembolism		26 (0.7)	17 (0.7)	9 (0.7)
Postoperative bleeding		47 (1.2)	28 (1.1)	19 (1.5)
Luxation		32 (0.8)	31 (1.2)	1 (0.1)
Delirium		24 (0.6)	20 (0.8)	4 (0.3)
Nerve damage		24 (0.6)	21 (0.8)	3 (0.2)

ASA, American Society of Anesthesiologists; BMI, body mass index.

of 0.072 and 0.086, respectively. The prediction model for delirium showed good discriminative value (AUC=85.9) and explained fraction of variance of 0.193. The models for luxation and VTE showed least favourable results on discrimination (AUC=58.4 and 66.3, respectively) and explained fraction of variance of 0.010 and 0.047, respectively.

The receiver operating characteristic curves and calibration plots for secondary outcomes are presented in online supplemental eFigure 1.

DISCUSSION

The prediction models developed in this study are aimed for personalised counselling and SDM in orthopaedic outpatient clinics. With our models, risk for SSI, VTE, POB, luxation, delirium and NER can be predicted by patient characteristics, comorbidities and medication use. For SSI, predicted probabilities range between 0.01% and 51.0%, which makes the model useful in adding relevant

personalised information for adequate SDM compared with the previously used population-based probability of risk of 3%.²⁶ However, it is important to state that the model showed moderately accurate prediction, especially when the risk is low. The model underestimates the risk with a predicted probability >10%. Therefore, predicted probabilities exceeding 10% should be interpreted with caution. Furthermore, other performance measures were moderate-to-reasonable, indicating moderate overall performance of the model for SSI. We found similar results for other outcomes, except for the model for luxation; this model seriously underestimates the risk for luxation and could therefore not be used for personalised counselling.

Our results are comparable with the results of a recent meta-analysis on impact of comorbidities on SSI in THA or TKA. The authors stated diabetes and liver disease to contribute to a higher risk for SSI.³ Another study with similar discriminative capacity found BMI, use of



Table 2 Models including the coefficient per predictor per surgical outcome

Variable	Surgical site infection		Venous thromboembolism		Postoperative bleeding		Luxation		Delirium		Nerve damage	
	Coefficient*	OR (95% CI)	Coefficient*	OR (95% CI)	Coefficient*	OR (95% CI)	Coefficient*	OR (95% CI)	Coefficient*	OR (95% CI)	Coefficient*	OR (95% CI)
Intercept	-7.272	-	-4.790	-	-7.172	-	-5.864	-	-14.307	-	-2.250	-
Age (years)	0.031	1.032 (1.005 to 1.059)	-0.008	0.991 (0.966 to 1.018)	0.033	1.034 (1.006 to 1.062)	0.013	1.014 (0.991 to 1.038)	0.127	1.137 (1.067 to 1.212)	-0.051	0.949 (0.926 to 0.974)
Gender (male/female)	-	-	-0.168	0.844 (0.377 to 1.888)	-	-	-	-	-	-	-0.254	0.772 (0.319 to 1.868)
BMI (kg/m ²)	-0.002	0.998 (0.937 to 1.063)	-	-	0.012	1.012 (0.954 to 1.073)	0.021	1.023 (0.951 to 1.099)	-	-	-	-
Obesity (yes/no)	-	-	1.376	4.040 (1.462 to 11.159)	-	-	-	-	-	-	-	-
Smoking status (yes/no)	0.757	2.145 (0.883 to 5.213)	-	-	-0.023	0.952 (0.336 to 2.701)	0.491	1.667 (0.651 to 4.268)	-	-	0.572	1.754 (0.510 to 6.029)
Lung disease (yes/no)	-	-	-	-	-	-	-	-	-	-	-	-
Immunological disorder (yes/no)	0.891	2.474 (1.186 to 5.158)	-	-	-	-	-	-	-	-	-	-
Rheumatoid arthritis (yes/no)	-	-	-	-	-	-	0.538	1.752 (0.408 to 7.530)	-	-	-	-
Diabetes mellitus (yes/no)	0.904	2.494 (1.125 to 5.529)	0.829	2.317 (0.870 to 6.173)	-	-	-	-	-	-	-	-
Liver disease (yes/no)	2.345	10.659 (2.441 to 46.555)	-	-	-	-	-	-	-	-	-	-
Heart disease (yes/no)	-	-	-	-	0.729	2.086 (1.040 to 4.183)	-	-	0.348	1.422 (0.590 to 3.428)	-	-
Disease of central nervous system (yes/no)	-	-	-	-	-	-	0.106	1.113 (0.324 to 3.822)	0.898	2.465 (0.936 to 6.490)	-	-
Thromboembolic event (yes/no)	-	-	1.501	4.586 (1.521 to 13.826)	-	-	-	-	-	-	-	-
Dysplasia (yes/no)	-	-	-	-	-	-	-	-	-	-	-0.009	0.993 (0.217 to 4.552)
Vitamin K antagonist use (yes/no)	-	-	-	-	0.787	2.220 (1.022 to 4.821)	-	-	-	-	-	-
NSAIDs (yes/no)	0.619	1.877 (0.946 to 3.725)	-	-	-	-	-	-	-	-	-	-

To calculate the absolute risk for the surgical complications: $P_{(surgical\ complication)} = 1/(1 + \exp(-linear\ part) \times 100\%$. Linear part=intercept + (coefficients x variables).
 *Adjustment for overfitting by shrinkage factor; the intercept was re-estimated.
 BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs.

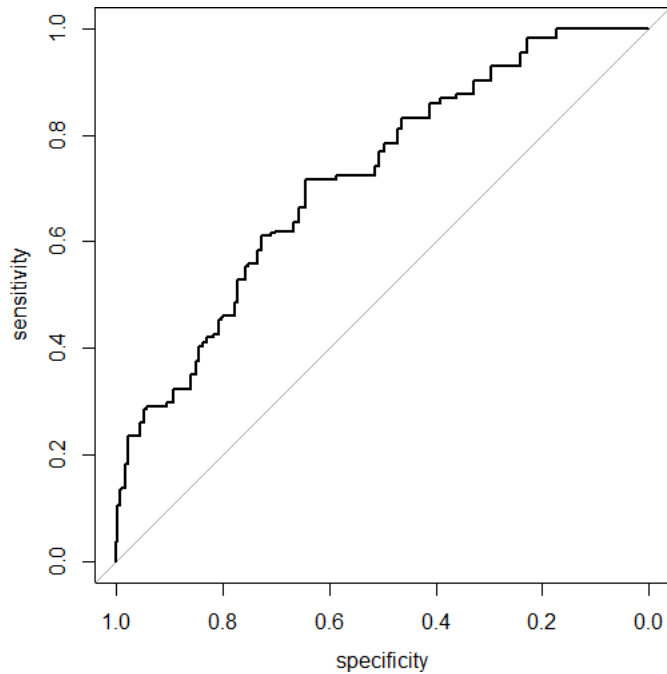


Figure 2 Receiver operating characteristic curve of the prediction model for surgical site infection area under the curve=71.9% (95% CI=69.4% to 74.4%).

immunosuppression, ASA score, procedure duration and prior surgeries as risk factors for SSI.³³ Some of these predictors did not contribute to a higher performance in our model and were therefore not included. We additionally found age to be a significant predictor for SSI. For the already available prediction model based on data of veterans with osteoarthritis of Harris *et al*, independent

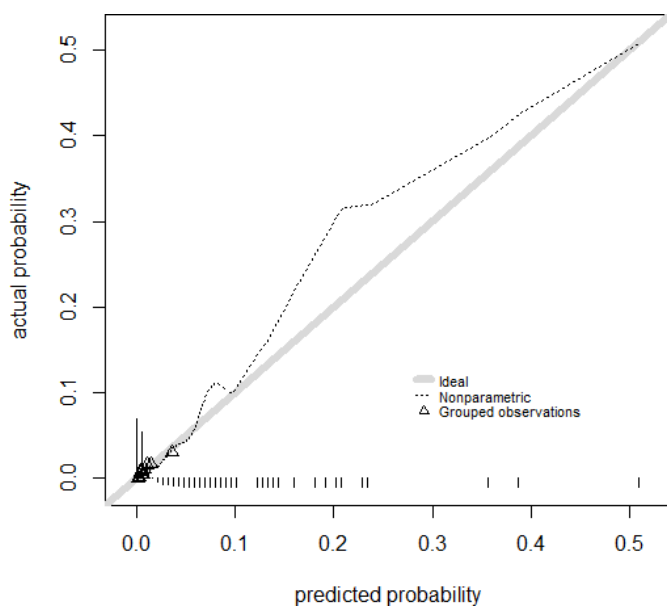


Figure 3 Calibration plot with the actual probability against the predicted probability for the model for surgical site infection. The triangles indicate quantiles ($g=10$) of patients with a similar predicted probability of success. The grey diagonal line represents perfect agreement between predicted and actual probability.

variables of the model cannot be compared for SSI since these results have not been reported.¹⁵ We found a slightly better c-statistic (AUC) of 0.72 compared with 0.66 in their boosted model. Similar variables as those used in our models were used for the development of other models predicting postoperative complications as well, such as the models of Harris *et al*. Unfortunately, a direct comparison of the predictive capacity of these variables between the models of Harris *et al* and our models is not possible, as the postoperative outcomes used in their prediction models were different to the postoperative outcomes used in our models.¹⁷ Also comparison with Bozic *et al*, is difficult since applicability to non-Medicare population is questionable, as they also describe in their discussion.¹⁶

Based on literature we expected use of thromboprophylaxis, such as platelet aggregation inhibitors, direct oral anticoagulants, low-molecular-weight heparin and/or vitamin K antagonists to be important predictors for POB. However, we could not demonstrate this finding in our model.³⁴ This is perhaps due to low frequencies of these predictors in our participants with POB and due to improved preoperative care regarding anticoagulant therapy. Our model for delirium included comparable predictors as other studies; they showed that age and pre-existing cognitive impairment are important predictors for delirium.^{35 36} Our model confirms this finding. Kalisvaart *et al*, developed a comparable model based on acute and elective hip surgery patients and found comparable predictors. The authors additionally found acute admission as predictor for delirium.³⁵ We cannot confirm this in our model since we focused on primary THA and TKA and these interventions are not primarily preferred in acute admissions due to hip fracture. The AUC indicates that our model is more accurate in estimating the risk for delirium (85.9 vs 73).³⁵

For VTE we only found obesity and thromboembolic event as significant risk factors.^{3 37} This can be explained by the fact that the recurrence rate is high after earlier thromboembolic events.³⁸ We could not demonstrate diabetes to be a significant predictor for VTE.³ For the risk of luxation, it is known that causes of dislocation are multifactorial and also caused by non-patient modifiable factors such as implant-related, surgery-related and hospital-related factors. It is unclear to what extent these factors contribute to the occurrence of luxation, but we expect these factors to be of influence the model.^{39 40} For these reasons, and the poor performance of the model for luxation, we consider this model of insufficient quality for use in patient information documents. Since we aimed our models to support preoperative SDM, we only used patient-related variables as these variables are considered modifiable.^{39 41}

Strengths and limitations

A strong point is that we thoroughly created a big data set and we used state-of-the-art statistics for our analyses. Furthermore, the simplicity of our models is a strength

**Table 3** Model performance

	Surgical site infection	Venous thromboembolism	Postoperative bleeding	Luxation	Delirium	Nerve damage
Brier score	0.010	0.007	0.012	0.012	0.006	0.008
Brier _{scaled}	0.026	0.007	0.010	0.003	0.027	0.012
Nagelkerke's R ²	0.081	0.047	0.072	0.010	0.193	0.086
AUC (95% CI)	71.9 (69.4 to 74.4)	66.3 (62.7 to 69.9)	73.0 (70.7 to 75.4)	58.4 (55.0 to 61.8)	85.9 (83.8 to 87.9)	76.6 (73.2 to 80.0)
H-L statistic (p value)	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001	p<0.001
Predicted possibilities						
Mean	0.010	0.007	0.012	0.012	<0.001	0.008
SD	0.015	0.007	0.012	0.004	0.012	0.010
Range	0.001–0.510	0.003–0.147	0.001–0.090	0.005–0.045	<0.001–0.147	0.001–0.072
Shrinkage factor	0.984	0.986	0.989	0.941	0.993	0.987

AUC, area under the curve; H-L, Hosmer and Lemeshow .

because we used predictors collected in usual care. The predictors are easy to assess and thereby easy to implement in care. Several limitations in this study should be noted. We retrospectively analysed prospectively collected data. These data were not primarily registered for research purposes and therefore their detail and accuracy could be less than optimal. Moreover, changes in reporting systems took place during the studied period, for instance the introduction of electronic medical records. It is known that changes in coding practice may change completeness of data.^{42 43} Although researchers performed data collection thoroughly, data about comorbidities and medication use could be missed because it was reported elsewhere. Moreover, we expect a small quantity of under-reporting regarding comorbidities since physicians and anaesthesiologists perchance make a selection of important comorbidities in their report. We tried to correct for this limitation by including medication use since all drugs are registered in preoperative anaesthesia report. Also, data from 2004 until 2018 were used. In this period preoperative care has been changed. To evaluate the effect of this change on our outcome, we checked our patterns of complications and found no differences in this period. Furthermore, due to a low estimated event rate (1%–3%) we needed a large population to have enough events to include predictors into our models. However, since not all predictors were significant in our final models, we expect that inclusion of more predictors would not lead to a considerably different model, as also discussed above. The models were developed based on pooled THA and TKA data. It is expected that the influence of patient characteristics, comorbidities and medication use is comparable for both THA and TKA.⁴⁴ The influence of comorbidities on outcomes is studied together quite often.³ Furthermore, we tested this assumption by performing the analysis on THA and TKA data only. The models with corresponding performance measures were still consistent with the main analysis.

Another limitation is that we only performed internal validation by bootstrapping, and were not yet able to determine external validity and clinical impact of the models. For clinical impact it is also important to determine the minimal clinically important difference of the outcomes.

CONCLUSION

Clinical prediction models were developed to contribute to more unbiased and accurate counselling in considering THA or TKA and are expected to be useful for identifying patients at risk for surgical complications. For SSI, the discriminative ability was reasonable and predicted risk varied between 0.01% and 51.0%. We expect the individual predicted risk to enhance SDM and support a well-founded choice. We consider our models for SSI, delirium and NER appropriate for clinical use when taking underestimation and overestimation of predicted risk into account. For clinical use of the models VTE and POB, caution concerning overestimation exceeding predicted probability of 0.08 and 0.05 (data presented in calibration plots in online supplemental eFigure 1), respectively, should be taken into account. Future studies should evaluate clinical impact and whether our models are feasible in an external population.

Supplementary information

In the online supplemental file, an Excel file with the prediction models calculator is provided, see online supplemental appendix 2. The decision aid including the prediction models is published in Dutch at the website of the Radboud university medical center.

Author affiliations

¹Radboud Institute of Health Sciences, Department of Orthopaedics, Radboud university medical center, Nijmegen, The Netherlands

²Radboud Institute of Health Sciences, IQ healthcare, Radboud university medical center, Nijmegen, The Netherlands

³Radboud Institute for Health Sciences, Department of Rehabilitation, Radboud university medical center, Nijmegen, The Netherlands

Acknowledgements We thank Anouk Tulp for her contribution to the data collection. Furthermore we would like to acknowledge Stefan Riemens for offering help in reviewing the manuscript for grammar and wording.

Contributors All authors confirm authorship on all four ICMJE criteria. Conceptualisation: LS, TJH, PvdW, SAWvdG. Data curation: LS. Formal analysis: LS, TJH, TvW, SAWvdG. Funding acquisition: SAWvdG. Investigation: LS, TvW. Methodology: LS, TJH, PvdW, SAWvdG. Project administration: LS. Resources: LS, TvW. Software—Supervision: TJH, PvdW, SAWvdG. Validation: LS. Visualisation: LS, TvW. Writing—original draft: LS. Writing—review and editing: TJH, TvW, PvdW, SAWvdG. Guarantor: LS.

Funding This work was supported by the Dutch National Healthcare Institute (Transparency about the quality of care 2018: using outcome information for shared decision-making). Grant number: N/A. The funding source was not involved during the study.

Competing interests PvdW participates in the Scientific Advisory Panel of the American Physical Therapy Association.

Patient and public involvement Patients and/or the public were involved in the design, or conduct, or reporting, or dissemination plans of this research. Refer to the Methods section for further details.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Medical Ethical Committee of Radboudumc. Reference number 2018-4880. The Institutional Review Board approval can be provided upon request. Anonymised data were used. If participants objected against use of their personal data, these data were not available for use in our analyses.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iD

Lieke Sweerts <http://orcid.org/0000-0001-6940-3367>

REFERENCES

- Smink AJ, van den Ende CHM, Vliet Vlieland TPM, *et al.* “Beating osteoARthritis”: Development of a stepped care strategy to optimize utilization and timing of non-surgical treatment modalities for patients with hip or knee osteoarthritis. *Clin Rheumatol* 2011;30:1623–9.
- Mota REM, Tarricone R, Ciani O, *et al.* Determinants of demand for total hip and knee arthroplasty: a systematic literature review. *BMC Health Serv Res* 2012;12:225.
- Podmore B, Hutchings A, van der Meulen J, *et al.* Impact of comorbid conditions on outcomes of hip and knee replacement surgery: a systematic review and meta-analysis. *BMJ Open* 2018;8:e021784.
- Hawker GA, Badley EM, Borkhoff CM, *et al.* Which patients are most likely to benefit from total joint arthroplasty? *Arthritis Rheum* 2013;65:1243–52.
- Ten Have IA, van den Bekerom MP, van Deurzen DF, *et al.* Role of decision AIDS in orthopaedic surgery. *World J Orthop* 2015;6:864–6.
- Chawla NV, Davis DA. Bringing big data to personalized healthcare: a patient-centered framework. *J Gen Intern Med* 2013;28 Suppl 3:660–5.
- Elwyn G. Idealistic, impractical, impossible? shared decision making in the real world. *Br J Gen Pract* 2006;56:403–4.
- Elwyn G, Edwards A, Hood K, *et al.* Achieving involvement: process outcomes from a cluster randomized trial of shared decision making skill development and use of risk communication AIDS in general practice. *Fam Pract* 2004;21:337–46.
- Meguid RA, Bronsert MR, Juarez-Colunga E, *et al.* Surgical risk preoperative assessment system (SURPAS): III. accurate preoperative prediction of 8 adverse outcomes using 8 predictor variables. *Ann Surg* 2016;264:23–31.
- Trickey AW, Ding Q, Harris AHS. How accurate are the surgical risk preoperative assessment system (SURPAS) universal calculators in total joint arthroplasty? *Clin Orthop Relat Res* 2020;478:241–51.
- Riddle DL, Golladay GJ, Jiranek WA, *et al.* External validation of a prognostic model for predicting nonresponse following knee arthroplasty. *J Arthroplasty* 2017;32:1153–8.
- Barlow T, Dunbar M, Sprowson A, *et al.* Development of an outcome prediction tool for patients considering a total knee replacement—the Knee Outcome Prediction Study (KOPS). *BMC Musculoskelet Disord* 2014;15:451.
- Lungu E, Desmeules F, Dionne CE, *et al.* Prediction of poor outcomes six months following total knee arthroplasty in patients awaiting surgery. *BMC Musculoskelet Disord* 2014;15:299.
- Tolk JJ, Waarsing JEH, Janssen RPA, *et al.* Development of preoperative prediction models for pain and functional outcome after total knee arthroplasty using the Dutch arthroplasty register data. *J Arthroplasty* 2020;35:690–8.
- Harris AH, Kuo AC, Bowe T, *et al.* Prediction models for 30-day mortality and complications after total knee and hip arthroplasties for veteran health administration patients with osteoarthritis. *J Arthroplasty* 2018;33:1539–45.
- Bozic KJ, Ong K, Lau E, *et al.* Estimating risk in Medicare patients with THA: an electronic risk calculator for periprosthetic joint infection and mortality. *Clin Orthop Relat Res* 2013;471:574–83.
- Harris AHS, Kuo AC, Weng Y, *et al.* Can machine learning methods produce accurate and easy-to-use prediction models of 30-day complications and mortality after knee or hip arthroplasty? *Clin Orthop Relat Res* 2019;477:452–60.
- Moons KGM, Altman DG, Reitsma JB, *et al.* Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–73.
- Castor EDC. Castor electronic data capture. Available: <https://castoredc.com> [Accessed 28 Aug 2019].
- van Steenberg LN, Denissen GAW, Spooren A, *et al.* More than 95% completeness of reported procedures in the population-based Dutch Arthroplasty Register. *Acta Orthop* 2015;86:498–505.
- de Waal Malefijt MC. An orthopedic complication-registration system. *Acta Orthop Scand* 1995;66:84–9.
- van der Wees PJ, Wammes JJG, Akkermans RP, *et al.* Patient-reported health outcomes after total hip and knee surgery in a Dutch university hospital setting: results of twenty years clinical Registry. *BMC Musculoskelet Disord* 2017;18:97.
- ZorginstituutNederland. Farmacotherapeutisch Kompas, 2020. Available: <https://farmacotherapeutischkompas.nl>
- Peduzzi P, Concato J, Kemper E, *et al.* A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–9.
- Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and COX regression. *Am J Epidemiol* 2007;165:710–8.
- van Arkel E, van der Kraan J, Hageman M. *Consultkaart Artrose in de heup*, 2016.
- Donders ART, van der Heijden GJMG, Stijnen T, *et al.* Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol* 2006;59:1087–91.
- Steyerberg EW. *Clinical prediction models*. New York: Springer, 2009.
- Salkind NJ. *Encyclopedia of measurement and statistics*, 2007.
- RUBIN DB. Inference and missing data. *Biometrika* 1976;63:581–92.
- Steyerberg EW, Vickers AJ, Cook NR, *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;21:128–38.
- Hosmer DW, Lemeshow S, Sturdivant R. Applied Logistic Regression. In: *Wiley series in probability and statistics*. 3rd edn. New York: Wiley, 2013.
- Berbari EF, Osmon DR, Lahr B, *et al.* The Mayo prosthetic joint infection risk score: implication for surgical site infection reporting and risk stratification. *Infect Control Hosp Epidemiol* 2012;33:774–81.
- Lindquist DE, Stewart DW, Brewster A, *et al.* Comparison of postoperative bleeding in total hip and knee arthroplasty patients



- receiving rivaroxaban, enoxaparin, or aspirin for thromboprophylaxis. *Clinical and Applied Thrombosis/Hemostasis* 2018;24:1315–21.
- 35 Kalisvaart KJ, Vreeswijk R, de Jonghe JFM, *et al.* Risk factors and prediction of postoperative delirium in elderly hip-surgery patients: implementation and validation of a medical risk factor model. *J Am Geriatr Soc* 2006;54:817–22.
- 36 Lindroth H, Bratzke L, Purvis S, *et al.* Systematic review of prediction models for delirium in the older adult inpatient. *BMJ Open* 2018;8, :e019223.
- 37 Migita K, Bito S, Nakamura M, *et al.* Venous thromboembolism after total joint arthroplasty: results from a Japanese multicenter cohort study. *Arthritis Res Ther* 2014;16:R154.
- 38 Hansson PO, Sörbo J, Eriksson H. Recurrent venous thromboembolism after deep vein thrombosis: incidence and risk factors. *Arch Intern Med* 2000;160:769–74.
- 39 Dargel J, Oppermann J, Brüggemann G-P, *et al.* Dislocation following total hip replacement. *Dtsch Arztebl Int* 2014;111:884–90.
- 40 Kunutsor SK, Barrett MC, Beswick AD, *et al.* Risk factors for dislocation after primary total hip replacement: meta-analysis of 125 studies involving approximately five million hip replacements. *Lancet Rheumatol* 2019;1:e1111–21.
- 41 Fessy MH, Putman S, Viste A, *et al.* What are the risk factors for dislocation in primary total hip arthroplasty? A multicenter case-control study of 128 unstable and 438 stable hips. *Orthop Traumatol Surg Res* 2017;103:663–8.
- 42 Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323–37.
- 43 Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *CMAJ* 2016;188:E158–64.
- 44 Dobson PF, Reed MR. Prevention of infection in primary THA and TKA. *EFORT Open Rev* 2020;5:604–13.