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

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

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.

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. This is remarkable, as discussing potential personal risks is an important aspect of SDM.

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. 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. 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. Another study developed a preoperative prediction model to predict residual reports on pain, functional outcome and treatment success for individual patients after TKA. Also useful electronic risk calculators predicting complications and mortality for patients and clinicians are available for specific populations. 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. 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. 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. 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, 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.

**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). 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. From the register of complications we extracted all surgeries and complications which occurred within 1 year after THA or TKA. In this registry, surgery-related orthopaedic complications were registered as well as other medical complications. All complications were registered by location code combined with a code for the nature of the complication. 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. 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. Medication use was reduced to the active substance of the drug and was categorised to drug groups according the Dutch pharmacotherapeutic compass.

**Sample size**
It is recommended that at least five events are collected for each predictor that is evaluated in multivariable regression analysis. 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%.

**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. 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. 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.

**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.

**Performance of the model**
We quantified measures of performance, discrimination and calibration. Overall model performance is the distance between predicted and actual outcome. To quantify overall model performance, we assessed Brier, Brier, 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 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. 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. 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.

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.

All statistical analyses were performed using R V.3.5.3. Packagesvim, mice, rms, pROC and generhoslem were used.
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$^2$
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.2 Another study with similar discriminative capacity found BMI, use of

<table>
<thead>
<tr>
<th>Table 1 Patient characteristics</th>
<th>Missing values</th>
<th>Total population (n=3776)</th>
<th>Total hip replacement (n=2494)</th>
<th>Total knee replacement (n=1282)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age, mean (SD), years</strong></td>
<td>0.10%</td>
<td>60.2 (15.8)</td>
<td>57.7 (17.0)</td>
<td>65.1 (11.7)</td>
</tr>
<tr>
<td><strong>Gender: female, no. (%)</strong></td>
<td>0.10%</td>
<td>2298 (60.9)</td>
<td>1468 (58.9)</td>
<td>829 (64.7)</td>
</tr>
<tr>
<td><strong>BMI, mean (SD), kg/m²</strong></td>
<td>2.60%</td>
<td>27.5 (5.2)</td>
<td>26.6 (4.7)</td>
<td>29.3 (5.6)</td>
</tr>
<tr>
<td><strong>Smoking: yes, no. (%)</strong></td>
<td>24.70%</td>
<td>498 (13.2)</td>
<td>341 (13.7)</td>
<td>157 (12.2)</td>
</tr>
<tr>
<td><strong>ASA classification, no. (%)</strong></td>
<td>0.40%</td>
<td>839 (22.2)</td>
<td>669 (26.8)</td>
<td>170 (13.3)</td>
</tr>
<tr>
<td><strong>Diagnosis hip, no. (%)</strong></td>
<td>0.40%</td>
<td>1599 (46.1)</td>
<td>1037 (40.9)</td>
<td>562 (43.7)</td>
</tr>
<tr>
<td><strong>Arthrosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rheumatoid arthritis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dysplasia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Osteonecrosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diagnosis knee, no. (%)</strong></td>
<td>0.90%</td>
<td>1915 (50.9)</td>
<td>1257 (50.4)</td>
<td>658 (51.3)</td>
</tr>
<tr>
<td><strong>Arthrosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rheumatoid arthritis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Side affected: right, no. (%)</strong></td>
<td>0.30%</td>
<td>1915 (50.9)</td>
<td>1257 (50.4)</td>
<td>658 (51.3)</td>
</tr>
<tr>
<td><strong>Surgical complications, no. (%)</strong></td>
<td>0%</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Surgical site infection</strong></td>
<td></td>
<td>38 (1.0)</td>
<td>25 (1.0)</td>
<td>13 (1.0)</td>
</tr>
<tr>
<td><strong>Venous thromboembolism</strong></td>
<td></td>
<td>26 (0.7)</td>
<td>17 (0.7)</td>
<td>9 (0.7)</td>
</tr>
<tr>
<td><strong>Postoperative bleeding</strong></td>
<td></td>
<td>47 (1.2)</td>
<td>28 (1.1)</td>
<td>19 (1.5)</td>
</tr>
<tr>
<td><strong>Luxation</strong></td>
<td></td>
<td>32 (0.8)</td>
<td>31 (1.2)</td>
<td>1 (0.1)</td>
</tr>
<tr>
<td><strong>Delirium</strong></td>
<td></td>
<td>24 (0.6)</td>
<td>20 (0.8)</td>
<td>4 (0.3)</td>
</tr>
<tr>
<td><strong>Nerve damage</strong></td>
<td></td>
<td>24 (0.6)</td>
<td>21 (0.8)</td>
<td>3 (0.2)</td>
</tr>
</tbody>
</table>

ASA, American Society of Anesthesiologists; BMI, body mass index.
Table 2  Models including the coefficient per predictor per surgical outcome

<table>
<thead>
<tr>
<th>Variable</th>
<th>Surgical site infection</th>
<th>Venous thromboembolism</th>
<th>Postoperative bleeding</th>
<th>Luxation</th>
<th>Delirium</th>
<th>Nerve damage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient* OR (95% CI)</td>
<td>Coefficient* OR (95% CI)</td>
<td>Coefficient* OR (95% CI)</td>
<td>Coefficient* OR (95% CI)</td>
<td>Coefficient* OR (95% CI)</td>
<td>Coefficient* OR (95% CI)</td>
</tr>
<tr>
<td>Intercept</td>
<td>-7.272 -</td>
<td>-4.790 -</td>
<td>-7.172 -</td>
<td>-5.864 -</td>
<td>-14.307 -</td>
<td>-2.250 -</td>
</tr>
<tr>
<td>Age (years)</td>
<td>0.031 1.032 (1.005 to 1.059)</td>
<td>-0.008 0.991 (0.966 to 1.018)</td>
<td>0.033 1.034 (1.006 to 1.062)</td>
<td>0.013 1.014 (0.991 to 1.038)</td>
<td>0.127 1.137 (1.067 to 1.212)</td>
<td>-0.051 0.949 (0.926 to 0.974)</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>- -</td>
<td>-0.168 -</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>-0.002 0.998 (0.937 to 1.063)</td>
<td>- -</td>
<td>0.012 1.012 (0.954 to 1.073)</td>
<td>0.021 1.023 (0.951 to 1.099)</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>Obesity (yes/no)</td>
<td>- -</td>
<td>1.376 4.040 (1.462 to 11.159)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>Smoking status (yes/no)</td>
<td>0.757 2.145 (0.883 to 5.213)</td>
<td>- -</td>
<td>-0.023 0.952 (0.336 to 2.701)</td>
<td>0.491 1.667 (0.651 to 4.268)</td>
<td>- -</td>
<td>0.572 1.754 (0.510 to 6.029)</td>
</tr>
<tr>
<td>Lung disease (yes/no)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>Immunological disorder (yes/no)</td>
<td>0.891 2.474 (1.186 to 5.158)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>Rheumatoid arthritis (yes/no)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>Diabetes mellitus (yes/no)</td>
<td>0.904 2.494 (1.125 to 5.529)</td>
<td>0.829 2.317 (0.870 to 6.173)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>Liver disease (yes/no)</td>
<td>2.345 10.659 (2.441 to 46.555)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>Heart disease (yes/no)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>0.729 2.086 (1.040 to 4.193)</td>
<td>- -</td>
<td>0.348 1.422 (0.590 to 3.428)</td>
</tr>
<tr>
<td>Disease of central nervous system (yes/no)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>0.729 2.086 (1.040 to 4.193)</td>
<td>- -</td>
<td>0.348 1.422 (0.590 to 3.428)</td>
</tr>
<tr>
<td>Thromboembolic event (yes/no)</td>
<td>- -</td>
<td>1.501 4.586 (1.521 to 13.826)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>Dysplasia (yes/no)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>Vitamin K antagonist use (yes/no)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>0.787 2.220 (1.022 to 4.821)</td>
<td>- -</td>
<td>-</td>
</tr>
<tr>
<td>NSAIDs (yes/no)</td>
<td>0.619 1.877 (0.946 to 3.725)</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>- -</td>
<td>-</td>
</tr>
</tbody>
</table>

To calculate the absolute risk for the surgical complications: \( P(surgical\ complication) = \frac{1}{1 + \exp(\text{linear part})} \times 100\% \). Linear part = intercept + (coefficients \times variables).

*Adjustment for overfitting by shrinkage factor; the intercept was re-estimated.

BMI, body mass index; NSAIDs, non-steroidal anti-inflammatory drugs.
immunosuppression, ASA score, procedure duration and prior surgeries as risk factors for SSI. 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 variables of the model cannot be compared for SSI since these results have not been reported. 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. Also comparison with Bozic et al, is difficult since applicability to non-Medicare population is questionable, as they also describe in their discussion.

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. 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. 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. 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).

For VTE we only found obesity and thromboembolic event as significant risk factors. This can be explained by the fact that the recurrence rate is high after earlier thromboembolic events. We could not demonstrate diabetes to be a significant predictor for VTE. 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. 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.

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.

---

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%).

Figure 3 Calibration plot with the actual probability against the predicted probability for the model for surgical site infection. The triangles indicate quantiles (q=10) of patients with a similar predicted probability of success. The grey diagonal line represents perfect agreement between predicted and actual probability.
because we used predictors collected in usual care. The
predictors are easy to assess and thereby easy to imple-
ment 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 accu-
racity 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 comor-
bidities 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 med-
ication 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 popu-
lation to have enough events to include predictors into
our models. However, since not all predictors were signifi-
cant 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 devel-
oped based on pooled THA and TKA data. It is expected
that the influence of patient characteristics, comorbi-
dities and medication use is comparable for both THA and
TKA.44 The influence of comorbidities on outcomes is
studied together quite often.7 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 vali-
dation by bootstrapping, and were not yet able to deter-
mine 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 consid-
ering THA or TKA and are expected to be useful for iden-
tifying 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

---

**Table 3 Model performance**

<table>
<thead>
<tr>
<th></th>
<th>Surgical site infection</th>
<th>Venous thromboembolism</th>
<th>Postoperative bleeding</th>
<th>Luxation</th>
<th>Delirium</th>
<th>Nerve damage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brier score</td>
<td>0.010</td>
<td>0.007</td>
<td>0.012</td>
<td>0.012</td>
<td>0.006</td>
<td>0.008</td>
</tr>
<tr>
<td>Brier scaled</td>
<td>0.026</td>
<td>0.007</td>
<td>0.010</td>
<td>0.003</td>
<td>0.027</td>
<td>0.012</td>
</tr>
<tr>
<td>Nagelkerke's $R^2$</td>
<td>0.081</td>
<td>0.047</td>
<td>0.072</td>
<td>0.010</td>
<td>0.193</td>
<td>0.086</td>
</tr>
<tr>
<td>AUC (95% CI)</td>
<td>71.9 (68.4 to 74.4)</td>
<td>66.3 (62.7 to 69.9)</td>
<td>73.0 (70.7 to 75.4)</td>
<td>58.4 (55.0 to 61.8)</td>
<td>85.9 (83.8 to 87.9)</td>
<td>76.6 (73.2 to 80.0)</td>
</tr>
<tr>
<td>H-L statistic (p value)</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>Predicted possibilities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>0.010</td>
<td>0.007</td>
<td>0.012</td>
<td>0.012</td>
<td>&lt;0.001</td>
<td>0.008</td>
</tr>
<tr>
<td>SD</td>
<td>0.015</td>
<td>0.007</td>
<td>0.012</td>
<td>0.004</td>
<td>0.012</td>
<td>0.010</td>
</tr>
<tr>
<td>Range</td>
<td>0.001–0.510</td>
<td>0.003–0.147</td>
<td>0.001–0.090</td>
<td>0.005–0.045</td>
<td>&lt;0.001–0.147</td>
<td>0.001–0.072</td>
</tr>
<tr>
<td>Shrinkage factor</td>
<td>0.984</td>
<td>0.986</td>
<td>0.989</td>
<td>0.941</td>
<td>0.993</td>
<td>0.987</td>
</tr>
</tbody>
</table>

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

Patient and public involvement
Patients and/or the public were involved in the design, 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).

REFERENCES
34 Lindquist DE, Stewart DW, Brewster A, et al. Comparison of postoperative bleeding in total hip and knee arthroplasty patients


