ACR Open Rheumatology
Vol. 0, No. 0, Month 2024, pp 1–9
DOI 10.1002/acr2.11709

AMERICAN COLLEGE of RHEUMATOLOGY
Empowering Rheumatology Professionals

Prediction of Early Adverse Events After THA: A Comparison of Different Machine-Learning Strategies Based on 262,356 Observations From the Nordic Arthroplasty Register Association (NARA) Dataset

Mikko S. Venäläinen,[1] Valtteri J. Panula,[2] Antti P. Eskelinen,[3] Anne Marie Fenstad,[4] Ove Furnes,[5] Geir Hallan,[5] Ola Rolfson,[6] Johan Kärrholm,[6] Nils P. Hailer,[7] Alma B. Pedersen,[8] Søren Overgaard,[9] Keijo T. Mäkelä,[10] and Laura L. Elo[11]

Objective. Preoperative risk prediction models can support shared decision-making before total hip arthroplasties (THAs). Here, we compare different machine-learning (ML) approaches to predict the six-month risk of adverse events following primary THA to obtain accurate yet simple-to-use risk prediction models.

Methods. We extracted data on primary THAs (N = 262,356) between 2010 and 2018 from the Nordic Arthroplasty Register Association dataset. We benchmarked a variety of ML algorithms in terms of the area under the receiver operating characteristic curve (AUROC) for predicting the risk of revision caused by periprosthetic joint infection (PJI), dislocation or periprosthetic fracture (PPF), and death. All models were internally validated against a randomly selected test cohort (one-third of the data) that was not used for training the models.

Results. The incidences of revisions because of PJI, dislocation, and PPF were 0.8%, 0.4%, and 0.3%, respectively, and the incidence of death was 1.2%. Overall, Lasso regression with stable iterative variable selection (SIVS) produced models using only four to five input variables but with AUROC comparable to more complex models using all 32 variables available. The SIVS-based Lasso models based on age, sex, preoperative diagnosis, bearing couple, fixation, and surgical approach predicted the risk of revisions caused by PJI, dislocations, and PPF, as well as death, with AUROCs of 0.61, 0.67, 0.76, and 0.86, respectively.

Conclusion. Our study demonstrates that satisfactory predictive potential for adverse events following THA can be reached with parsimonious modeling strategies. The SIVS-based Lasso models may serve as simple-to-use tools for clinical risk assessment in the future.

INTRODUCTION

Although primary total hip arthroplasty (THA) is a safe and efficient intervention, early adverse events, such as revision and death, occur. Approximately 2% to 3% of primary THAs are revised within the first postoperative year, with dislocation (12%–33%), periprosthetic joint infection (PJI; 11%–23%), and periprosthetic fracture (PPF; 5%–18%) being the most frequently

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Dr Elo's work was supported by the European Research Council (ERC) (grant 677943), European Union's Horizon 2020 research and innovation programme (grant 955321), the Academy of Finland (grants 310561, 314443, 329278, 335434, 335611, and 341342), and the Sigrid Juselius Foundation. Dr Venäläinen's work was supported by the Academy of Finland (grant 322123) and the state research funding of well-being services county of Southwest Finland. Dr Hailer's work was supported by the Swedish Research Council (VR 2021-00980).

[1]Mikko S. Venäläinen, PhD: Turku University Hospital, University of Turku and Åbo Akademi University, Turku, Finland; [2]Valtteri J. Panula, MD, PhD: Turku University Hospital and University of Turku, Turku, Finland; [3]Antti P. Eskelinen, MD, PhD: Coxa Hospital for Joint Replacement and University of Tampere, Tampere, Finland, and the Finnish Arthroplasty Register, Finnish Institute for Health and Welfare, Helsinki, Finland; [4]Anne Marie Fenstad, MSc: Haukeland University Hospital, Bergen, Norway; [5]Ove Furnes, MD, PhD, Geir Hallan, MD, PhD: Haukeland University Hospital and University of Bergen, Bergen, Norway; [6]Ola Rolfson, MD, PhD, Johan Kärrholm, MD, PhD: University of Gothenburg, Gothenburg, Sweden; [7]Nils P. Hailer, MD: Uppsala University Hospital, Uppsala, Sweden; [8]Alma B. Pedersen, MD, PhD, DMSc: Aarhus University Hospital and Aarhus University, Aarhus, Denmark; [9]Søren Overgaard, MD, PhD: Copenhagen University Hospital and University of Copenhagen, Copenhagen, Denmark; [10]Keijo T. Mäkelä, MD, PhD: Turku University Hospital and University of Turku, Turku, Finland, and the Finnish Arthroplasty Register, Finnish Institute for Health and Welfare, Helsinki, Finland; [11]Laura L. Elo, PhD: University of Turku and Åbo Akademi University, Turku, Finland.

Additional supplementary information cited in this article can be found online in the Supporting Information section (http://onlinelibrary.wiley.com/doi/10.1002/acr2.11709).

Author disclosures are available at https://onlinelibrary.wiley.com/doi/10.1002/acr2.11709.

Address correspondence via email to Mikko Venäläinen, PhD, at mikko.venalainen@utu.fi; or to Laura L. Elo, PhD, at laura.elo@utu.fi.

Submitted for publication December 10, 2023; accepted in revised form June 8, 2024.

registered reasons.[1–4] Furthermore, despite the constantly reducing rates, postoperative mortality remains a recognized complication, especially among older adult patients.[5] Recently, incidence rates of approximately 0.9% to 1.2% have been reported at 1 year after the primary THA.[4,6,7]

With an aging and increasingly obese population, the incidence of primary THAs is expected to increase by up to 200% by 2030.[8–10] Simultaneously, primary THAs are increasingly being performed among more obese patients with more comorbidities who are also known to be at elevated risk of short-term complications.[1,11,12] Consequently, based on the recent trends, increases of 31% to 70% in the volumes of revision THA have been estimated in the United States, England, and Wales by 2030.[13,14] Overall, the estimated increases in both primary and revision THAs impose a substantial challenge to health care systems worldwide, and novel strategies for optimizing treatment outcomes and avoiding unnecessary complications are needed.

To alleviate the revision burden and minimize any unnecessary risks, directing the right type of treatment to the right individual and identifying high-risk patients requiring more intensive follow-up play a central role. To preoperatively evaluate the risk of revision and death following primary THA, several multivariable risk prediction models have been introduced.[15–20] Although the issue is global, these models have typically been developed and internally validated using data from a single arthroplasty register. Furthermore, the majority of the presented models rely on conventional regression modeling strategies[16] that might be outperformed by machine-learning (ML)–based approaches, especially when the event of interest is rare.[21] However, to achieve improved performance, ML methods require typically much more data than the conventional approaches,[22] and, hence, the best results are expected to be achieved when the amount of data, in terms of both the number of cases as well as variables included, is scaled up as high as possible.

In the present study, we applied a range of well-established ML algorithms with varying complexity to the Nordic Arthroplasty Register Association (NARA) dataset, a unified representation of the national hip arthroplasty registries of Sweden, Norway, Denmark, and Finland, to compare different modeling strategies for predicting the risk of the most common revision outcomes and death following primary THA. With the use of multinational data, we aimed at developing generalizable models that could be easily applied to evaluate preoperative risk estimates for an individual patient based on typical patient characteristics and planned surgical parameters in any modern health care setting.

PATIENTS AND METHODS

Study population. Initially, all primary THAs registered in the NARA dataset between 1995 and 2018 were extracted for analysis. The dataset consists of pooled data from the national hip arthroplasty registries of Sweden, Norway, Denmark, and Finland and has been described in more detail previously.[23,24] Ethical approval for the register-based study was granted by the appointed authority in each participating country: the Swedish Ethical Review Authority (1184-18/2019-00812), the Finnish National Institute of Health and Welfare (Dnro THL/1743/5.05.00/2014), the Norwegian Data Inspectorate (ref 24.1.2017: 16/01622-3/CDG), and the Danish Data protection agency (1-16-02-54-17).

The initial dataset consisted of 848,787 primary hip arthroplasties. Because of several changes in clinical practice regarding the use of implant materials, femoral head size, and fixation, we restricted our analyses to primary THAs performed since 2010, representing the most current clinical practice (Figure 1).

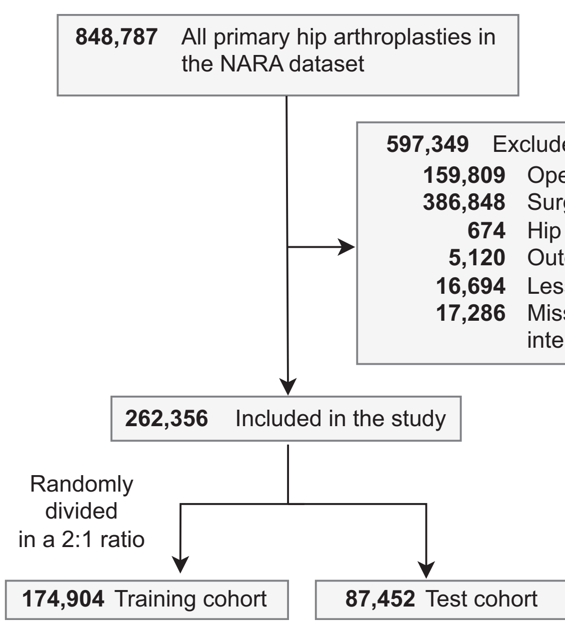

Figure 1. Selection of patients into the study. NARA, Nordic Arthroplasty Register Association.

Furthermore, we retained only primary THAs that had a minimum follow-up time of at least 6 months to exclude operations performed close to the last available date in the dataset for which the occurrence of primary outcomes could not be verified. To prevent dependent observations after bilateral arthroplasty, we included only the first operation reported for each patient. Finally, because some ML algorithms required the use of complete data, we included only patients with complete information for all candidate predictors to enable direct comparison of the predictions obtained using different methods. Because only 6% of the patients were excluded because of missing data, the excluded data points were assumed to be missing completely at random and not induce any bias in the present analyses.[25,26] Overall, this left us with a total of 262,356 primary THAs. Finally, the data were divided in a 2:1 ratio into separate training (random sample of 67% of the population, n = 174,904) and test (random sample of 33% of the population, n = 87,452) cohorts, used for developing and internally validating the models, respectively.

Study outcomes and candidate predictors. For each primary operation, we considered as our main outcomes of interest the first revision surgical procedure owing to the three most common reasons, PJI, dislocation or PPF, and death, during the first 6 postoperative months. During modeling, each of these outcomes was treated as a separate binary outcome for which specific risk prediction models were developed. Other reasons for revision were not considered for risk prediction modeling. In all countries, revision procedure was defined as a surgical procedure including the exchange or removal of any component(s). The candidate predictors considered for risk prediction models included previously identified risk factors for adverse events following THA. The considered patient characteristics included age,[19,27] sex,[19,28] simultaneous bilateral operation,[29] and primary diagnosis,[19,28,30] whereas surgical characteristics included fixation type,[31] the use of trochanteric osteotomy,[32] surgical approach (posterior or nonposterior, including anterior, anterolateral, and others),[19,28,30] bearing couple (recoded based on the combination of cup and caput materials),[29,33] the diameter of the femoral head,[19,28] and the presence of hydroxyapatite coating on the cup or stem.[34] The candidate predictors and other baseline information have been summarized in more detail in Table 1.

Model development and statistical analysis. Optimally, a prediction model suitable for clinical use should be both accurate and easy to use, using only data that are essential for the predictions. To identify the best modeling approach for each prediction task, we applied a range of ML algorithms to the primary THA data in the training cohort, namely logistic regression, classification tree modeling, random forest (RF), gradient-boosting machines (GBMs), penalized logistic regression with both Lasso penalty (Lasso regression) and Ridge penalty (Ridge regression), naive Bayes, and neural networks, which are among the most common and popular methods used widely for binary classification in various application areas.[35] Among the applied models, logistic regression is considered as the most conventional approach and can be used as a reference for the other ML algorithms. Here, the logistic regression models were trained using all candidate predictors without any additional variable selection. Finally, we also applied Lasso regression in combination with the stable iterative variable selection (SIVS) procedure previously suggested as an efficient method for developing simple-to-use risk prediction models with fewer variables but retaining the same discrimination performance as the more complex models.[19,36,37] The discrimination performances of the ML models were evaluated in the test cohort in terms of the area under the receiver operating characteristic curve (AUROC). For approximating model complexity, we determined the number of nonzero regression coefficients or variables and model-generated intervariable interactions with nonzero influence in top-performing algorithms. The calibration of predicted risks, that is, the agreement between the observed outcomes and predictions, was evaluated by grouping individuals by deciles of the predicted risk. Further information on the applied ML methods, their hyperparameters, and used software packages can be found in the Supplementary Material.

All statistical analyses and mathematical modeling were conducted using R statistical computing environment version 4.0.3 (R Core Team, 2016. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/). In addition to method-specific packages reported in the Supplementary Material (Supplementary Table S1), R packages *ggplot2*,[38] and *pROC*[39] were used for the visualization of results and evaluation of AUROC values, respectively. Comparisons between the characteristics of training and test cohorts were performed using the Mann-Whitney test for continuous variables and the chi-squared test for categorical variables. The level of significance in all statistical comparisons was set at $P < 0.05$.

RESULTS

Characteristics of the study population. The patients in the study cohort were, on average, aged 68 years, were typically female (59%), and had their hips operated for primary osteoarthritis (79%), mostly using uncemented fixation (42%) (Table 1). No statistically significant differences were observed between the characteristics of the training and test cohorts. Of the 262,356 hips included, within 6 months, 2,074 (0.8%) were revised because of PJI, 1,104 (0.4%) because of dislocation, and 759 (0.3%) because of PPF (Table 2). Other reasons for revision were registered for 756 (0.3%) hips. A total of 3,144 (1.2%) deaths occurred during the first 6 postoperative months.

Table 1. Patient and procedure characteristics for the included operations in the training and test cohorts. All other variables except for country and laterality were considered during predictive modeling*

Characteristics	Training cohort (n = 174,904)	Test cohort (n = 87,452)
Country, n (%)		
Denmark	39,715 (22.7)	19,653 (22.5)
Norway	30,335 (17.4)	15,229 (17.4)
Sweden	73,003 (41.7)	36,558 (41.8)
Finland	31,851 (18.2)	16,012 (18.3)
Age, mean (SD), y	68.1 (11.1)	68.1 (11.0)
Sex, n (%)		
Female	102,199 (58.4)	51,406 (58.8)
Male	72,705 (41.6)	36,046 (41.2)
Laterality, n (%)		
Right	98,995 (56.6)	49,635 (56.8)
Left	75,909 (43.4)	37,817 (43.2)
Simultaneous bilateral operation, n (%)		
No	173,455 (99.2)	86,723 (99.2)
Yes	1,449 (0.8)	729 (0.8)
Preoperative diagnosis, n (%)		
Primary osteoarthritis	138,661 (79.3)	69,331 (79.3)
Hip fracture	19,146 (10.9)	9,515 (10.9)
Nontraumatic femoral head necrosis	3,988 (2.3)	2,109 (2.4)
Rheumatoid arthritis	1,752 (1.0)	900 (1.0)
Ankylosing spondylitis	175 (0.1)	81 (0.1)
Developmental dysplasia of the hip	5,209 (3.0)	2,524 (2.9)
Slipped capital femoral epiphysis	206 (0.1)	84 (0.1)
Perthes disease	649 (0.4)	283 (0.3)
Combination of slipped capital femoral epiphysis and Perthes disease	78 (<0.1)	34 (<0.1)
Other inflammatory	495 (0.3)	227 (0.3)
Others	4,545 (2.6)	2,364 (2.7)
Fixation, n (%)		
Cemented	63,801 (36.5)	31,840 (36.4)
Hybrid	16,990 (9.7)	8,577 (9.8)
Inverse hybrid	20,423 (11.7)	10,323 (11.8)
Uncemented	73,690 (42.1)	36,712 (42.0)
Surgical approach, n (%)		
Anterior, anterolateral and others	77,873 (44.5)	38,779 (44.3)
Posterior	97,031 (55.5)	48,673 (55.7)
Bearing couple, n (%)		
CoC	9,030 (5.2)	4,372 (5.0)
CoX	26,723 (15.3)	13,570 (15.5)
CoP	3,688 (2.1)	1,943 (2.2)
MoP	25,475 (14.6)	12,666 (14.5)
MoX	108,615 (62.1)	54,247 (62.0)
Other	1,373 (0.7)	654 (0.8)
Hydroxyapatite coating (cup), n (%)		
No	146,844 (84.0)	73,633 (85.5)
Yes	28,060 (16.0)	13,819 (14.5)
Hydroxyapatite coating (stem), n (%)		
No	107,777 (61.6)	53,622 (61.3)
Yes	67,127 (38.4)	33,830 (38.7)
Caput size, n (%)		
22 mm	1,120 (0.6)	521 (0.6)
28 mm	31,712 (18.1)	16,127 (18.4)
32 mm	85,458 (48.9)	42,651 (48.8)
36 mm	54,543 (31.2)	27,160 (31.1)
>36 mm	1,991 (1.1)	968 (1.1)
Other	80 (0.1)	25 (<0.1)
Trochanteric osteotomy, n (%)		
No	174,582 (99.8)	87,269 (99.8)
Yes	322 (0.2)	183 (0.2)

(Continued)

Table 1. (*Cont'd*)

*CoC, ceramics on ceramics; CoP, ceramics on conventional (noncrosslinked) polyethylene; CoX, ceramics on polyethylene crosslink; MoP, metal on conventional (noncrosslinked) polyethylene; MoX, metal on polyethylene crosslink.

Comparison of the ML methods. Overall, GBM was the best-performing ML algorithm with AUROCs of 0.61 (95% confidence interval [CI] 0.59–0.63), 0.68 (95% CI 0.65–0.70), and 0.77 (95% CI 0.74–0.79) for revisions caused by PJI, dislocation, and PPF, respectively, as well as with an AUROC of 0.87 (95% CI 0.86–0.88) for death (Table 3). However, in terms of AUROCs, there were no substantial differences among the six top-performing models, including GBM, conventional logistic regression, Lasso regression, Ridge regression, Lasso regression with SIVS, and RF. Compared with GBM, the largest difference among these was a lower AUROC of RF (AUROC 0.85, 95% CI 0.84–0.86) for predicting death.

Among the top-performing models, Lasso regression with SIVS was able to produce models with the least complexity but with similar accuracy as the more complex models (Figure 2A). In total, the SIVS-based model for PJI included only four (13%) variables and models for dislocation, PPF, and death, only five (16%) variables each of the available 32 with nonzero influence on risk predictions. Among all input data, Lasso with SIVS identified age, sex, preoperative diagnosis, bearing couple, fixation, and surgical approach as the minimum set of variables sufficient for accurate risk predictions (Figure 2B). In contrast, all the competing methods used nearly all 11 available variable types and associated information to reach similar AUROCs. In all comparisons, logistic regression, conventional Lasso regression, and Ridge regression had virtually the same performance and variables. Because of the good performance with the minimum number of input variables, SIVS-based Lasso models were considered for further evaluation as practical risk prediction models.

Simple-to-use risk prediction models obtained using Lasso regression with SIVS. The variables and regression coefficients obtained for each outcome using Lasso regression with SIVS are summarized with example risk calculations in Table 4. Additional details on using the regression

Table 2. The rates of short-term revision outcomes and death following primary total hip arthroplasty in the study population

Outcome	All patients (N = 262,356)	Training cohort (n = 174,904)	Test cohort (n = 87,452)
Revision procedure, n (%)	4,767 (1.8)	3,133 (1.8)	1,634 (1.9)
Periprosthetic joint infection	2,074 (0.8)	1,373 (0.8)	701 (0.8)
Dislocation	1,104 (0.4)	696 (0.4)	408 (0.5)
Periprosthetic fracture	759 (0.3)	504 (0.3)	255 (0.3)
Aseptic loosening	298 (0.1)	202 (0.1)	96 (0.1)
Other	458 (0.2)	304 (0.2)	154 (0.2)
Reason missing	74 (<0.1)	54 (<0.1)	20 (<0.1)
Death, n (%)	3,144 (1.2)	2,091 (1.2)	1,053 (1.2)

Table 3. Discrimination performance of the applied machine-learning methods in terms of the AUROC in the independent test cohort*

Method	Periprosthetic joint infection AUROC (95% CI)	Dislocation AUROC (95% CI)	Periprosthetic fracture AUROC (95% CI)	Death AUROC (95% CI)
Gradient-boosting machines	0.61 (0.59–0.63)	0.68 (0.65–0.70)	0.77 (0.74–0.79)	0.87 (0.86–0.88)
Lasso regression	0.61 (0.59–0.63)	0.67 (0.64–0.69)	0.76 (0.73–0.79)	0.87 (0.85–0.88)
Lasso regression with SIVS	0.61 (0.59–0.63)	0.67 (0.64–0.69)	0.76 (0.73–0.78)	0.86 (0.85–0.87)
Logistic regression	0.61 (0.59–0.63)	0.67 (0.64–0.69)	0.76 (0.73–0.78)	0.87 (0.85–0.88)
Ridge regression	0.61 (0.59–0.63)	0.67 (0.64–0.69)	0.75 (0.73–0.78)	0.87 (0.85–0.88)
Random forest	0.61 (0.59–0.63)	0.68 (0.65–0.70)	0.77 (0.74–0.79)	0.85 (0.84–0.86)
Neural network	0.61 (0.59–0.63)	0.66 (0.64–0.69)	0.75 (0.72–0.78)	0.86 (0.85–0.87)
Naive Bayes	0.59 (0.57–0.61)	0.66 (0.64–0.69)	0.72 (0.70–0.75)	0.85 (0.83–0.86)
Classification tree	0.60 (0.58–0.62)	0.55 (0.53–0.57)	0.74 (0.71–0.76)	0.85 (0.84–0.86)

*AUROC, area under receiver operating characteristic curve; CI, confidence interval; SIVS, stable iterative variable selection.

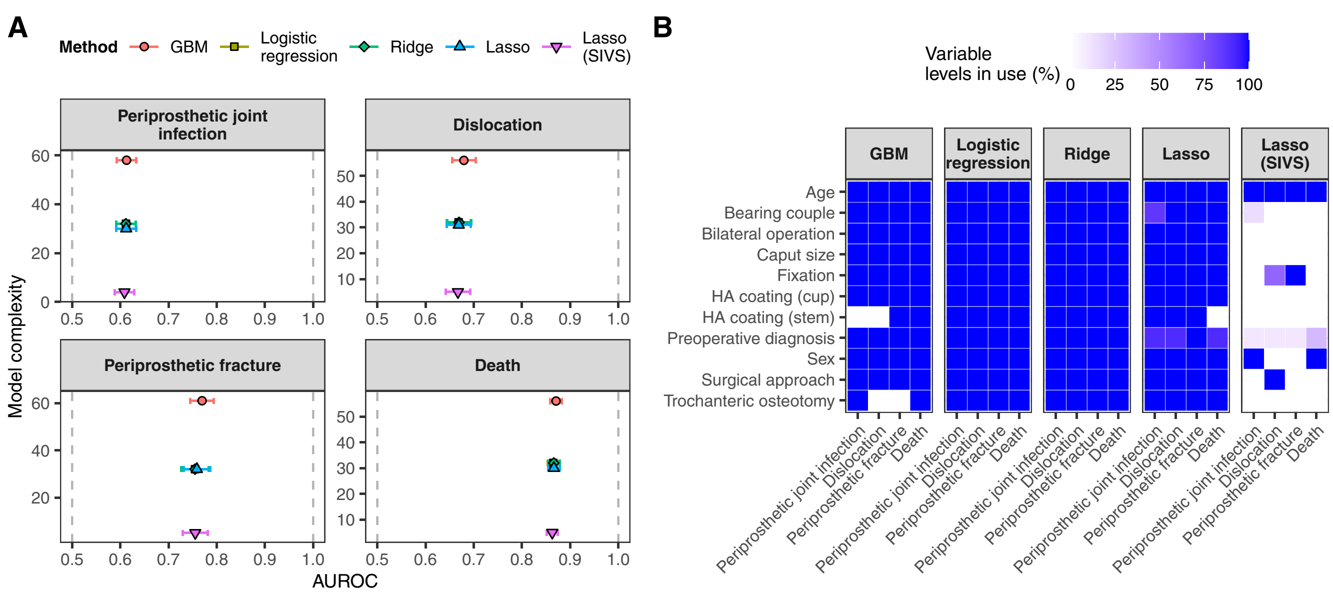

Figure 2. Evaluation of the variables used by the top-performing models. (A) The complexity of the models in terms of the number of regression coefficients or variables and intervariable interactions (specifically in GBMs) with nonzero influence on model predictions versus discrimination performance in terms of AUROC in the test cohort. Horizontal lines indicate 95% confidence intervals. Ridge regression, conventional Lasso regression, and logistic regression had nearly identical performance. (B) Summary of the variables with nonzero influence identified by different modeling approaches. The color indicates the fraction of variable levels with nonzero influence in the final models. AUROC, area under the receiver operating characteristic curve; GBM, gradient-boosting machine; HA, hydroxyapatite; SIVS, stable iterative variable selection.

Table 4. Regression coefficients[a] in the Lasso models built using stable iterative variable selection procedure*

Variable	Regression coefficient (β) for model: PJI	Dislocation	PPF	Death
Age (per 10 years)	0.132	0.164	0.353	0.802
Sex				
Female	–	–	–	–
Male	0.496	–	–	0.502
Preoperative diagnosis				
Primary osteoarthritis	–	–	–	–
Hip fracture	0.389	0.980	0.583	2.265
Nontraumatic femoral head necrosis	–	–	–	1.281
Rheumatoid arthritis	–	–	–	–
Others	–	–	–	2.777
Fixation				
Uncemented	–	0.801	2.606	–
Cemented	–	–	–	–
Hybrid	–	0.775	0.890	–
Inverse hybrid	–	–	2.039	–
Bearing				
MoX	–	–	–	–
MoP	–	–	–	–
CoX	–	–	–	–
CoC	−0.542	–	–	–
CoP	–	–	–	–
Other	–	–	–	–
Surgical approach				
Anterior, anterolateral, and others	–	–	–	–
Posterior	–	0.355	–	–
Example calculations[b]				
Raw patient score (sum of patient value × β coefficient)	1.379	2.565	3.231	8.280
Intercept	−6.013	−7.501	−10.214	−11.345
Transformed score = $\frac{1}{1+\exp(-(Intercept + Raw\ score))}$	0.010 or 1.0%	0.007 or 0.7%	0.001 or 0.1%	0.045 or 4.5%

*CoC, ceramics on ceramics; CoP, ceramics on conventional (noncrosslinked) polyethylene; CoX, ceramics on polyethylene crosslink; MoP, metal on conventional (noncrosslinked) polyethylene; MoX, metal on polyethylene crosslink; PJI, periprosthetic joint infection; PPF, periprosthetic fracture.

[a]The beta coefficients indicate the impact of one-unit change in a predictor variable, given in parentheses, on the response variable when the other predictors are held constant. A positive coefficient indicates risk-increasing effect and negative risk-decreasing effect. Fields without a numerical value indicate that the specific variable is not needed for predicting the risk of the designated outcome (ie, regression coefficient equals zero) and, therefore, for categorical variables, functions as a reference group.

[b]Example calculations of the average estimates of risk are given for 75-year-old female patients with hip fracture diagnosis and no simultaneous bilateral operation who are having a cemented total hip arthroplasty surgery performed using the posterior approach (no trochanteric osteotomy) and with implant components having a metal on polyethylene bearing surface, 36-mm femoral head size, and no hydroxyapatite coating on the cup. More details on the calculations and additional examples (Supplementary Tables S2-S5) can be found in the Supplementary Material.

coefficients for risk prediction and further examples (Supplementary Tables S2-S5) can be found in the Supplementary Material. For PJI, the model identified advanced age, male sex, and a preoperative diagnosis of hip fracture as risk factors, whereas ceramic-on-ceramic bearings decreased the risk compared with other bearing types. For dislocation, advanced age, a preoperative diagnosis of hip fracture, uncemented and hybrid fixations, and posterior approach were identified as risk factors. For PPF, the identified risk factors were advanced age; a preoperative diagnosis of hip fracture; and the use of uncemented, hybrid, or inverse hybrid fixations. For death, advanced age; male sex; and a preoperative diagnosis of hip fracture, nontraumatic femoral head necrosis, or other unspecified diagnosis were identified as key risk factors. All risk predictions made using the simple-to-use SIVS-based models were in good agreement with the observed outcome rates and showed no signs of substantial overfitting or underfitting (Supplementary Figure S1).

DISCUSSION

In the present study, we compared a range of ML algorithms to identify the best modeling approach for predicting the risk of the most common short-term revision outcomes (ie, PJI, dislocation, and PPF) as well as death within 6 months from the primary THA, based on the NARA dataset. We observed that there was little difference in the obtained AUROCs between the applied methods and that the complexity and number of required variables in risk prediction models can be greatly reduced with minimal loss in prediction accuracy. Finally, by using Lasso regression with SIVS, the modeling strategy requiring the fewest input variables, we developed simple-to-use preoperative risk prediction models that may assist in preoperative estimations of the expected levels of risks and clinical decision-making in the future.

A key finding in our study was that, despite the large amounts of operations in THA register data, accurate predictions can be obtained even with simpler modeling strategies. Similar benefits of Lasso regression in the reduction of input variables have also been reported before,[40] but here the Lasso regression accompanied with the SIVS procedure produced substantially simpler models without any reductions in prediction accuracy compared with the conventional Lasso. This implies that with careful variable selection, the most essential relationships with each outcome of interest can be captured with simple linear relationships and that models for revisions and death following THA based on registry data do not necessarily benefit from more sophisticated approaches involving modeling of deep intervariable interactions and complex nonlinear relationships. This observation is identical to our previous studies using the same approach in other prediction tasks.[36,37] The use of simpler modeling strategies is also more practical because the models can be applied using simple risk equations without dedicated computer software. Finally, the effect of each risk factor and the obtained results are also easier to interpret, helping to communicate the expectations of the operation with the patient.

Overall, the model for death within the first 6 postoperative months reached the highest discrimination performance and was comparable to the excellent performance observed in our previous risk prediction study using the Finnish Arthroplasty Register (FAR).[19] The Lasso regression with SIVS identified advanced age, male sex, preoperative hip fracture, nontraumatic femoral head necrosis, or other unspecified preoperative diagnosis as the most important variables increasing the risk of death. Similar findings concerning intuitive or well-established risk factors, such as advanced age,[27,41] male sex,[27,42] and hip fracture,[43] have thorough previous documentation.

In contrast to our previous study,[19] the model for revisions owing to PPF reached substantially better performance (NARA AUROC 0.76 vs FAR AUROC 0.65). Although the revision rates were quite similar between the two studies (0.3% vs 0.5%), the current dataset contained approximately 10 times more operations, including more cases with cemented stems, potentially explaining the improvement, because the ML algorithms had substantially more material for training the models. Again, the Lasso regression with SIVS also identified risk factors associated with revisions caused by PPF before, including advanced age; preoperative diagnosis of hip fracture; and the use of uncemented, hybrid, and inverse hybrid fixations.[31,44–46]

The model predicting the risk of dislocation reached similar moderate performance as in our previous study (NARA AUROC 0.67 vs FAR AUROC 0.65)[19] and consisted of several known risk factors, such as advanced age, preoperative hip fracture diagnosis, and posterior approach.[28,30] Furthermore, Thoen et al recently reported elevated dislocation risk after the use of uncemented fixation compared with cemented and inverse hybrid fixations,[47] thus supporting the selection of uncemented and hybrid fixations as risk factors. This finding, however, could also be explained by time-dependent confounding related to the increased use of uncemented fixation in the more recent time period.

The models for revisions because of PJI reached slightly lower performance compared with our previous study using the FAR data[19] (NARA AUROC 0.61 vs FAR AUROC 0.68) as well as the risk calculator developed based on the Swedish Arthroplasty Register (AUROC 0.68).[20] However, the model consisted of previously identified risk factors, such as male sex and preoperative hip fracture.[33] The ceramic-on-ceramic bearing has also previously been associated with a reduced infection revision risk,[29,33,48] although the finding might be affected by residual confounding because this bearing type tends to be used in younger and healthier patients with fewer comorbidities.

Even though all models reached moderate to excellent discrimination performance and the model predictions were in good agreement with the observed outcome rates, our study still has several limitations. First, the completeness of revision arthroplasties in the NARA member countries is in the range of 85% to 94%,[49] indicating that not all revisions are reported to the national registries, causing potential bias for our results. Second, the NARA dataset contains only the variables that all countries can deliver, and not all key risk factors for each outcome have been included during modeling. For example, greater body mass index and the American Society of Anesthesiologists physical status classification have previously been listed as important risk factors for revisions because of infection,[16,20,29,50] and thus their inclusion might have led to even simpler models and improved performance. Similarly, the model for death might be further simplified by replacing some of the variables with the American Society of Anesthesiologists physical status classification, a significant risk factor for mortality following THA.[27,51,52] Overall, the benefit of ML methods might become more apparent after the inclusion of more complex data and novel additional risk factors. Finally, regardless of large amounts of operations in training and test cohorts from four countries, it would be beneficial to externally validate the performance of the developed models in additional patient cohorts that could reveal the potential need for the recalibration of model coefficients and to identify potentially redundant variables.[20,52]

In conclusion, the present study demonstrates that when predicting revision and death within 6 months of primary THA based on arthroplasty register data, simpler models can achieve performance equal to that of complex modeling strategies but with reduced model complexity and improved usability. The simple-to-use and intuitive models developed using Lasso regression with SIVS for PJI, dislocation, PPF, and death all reached moderate to excellent performance. Once externally validated, the developed models have potential to facilitate clinical decision-making by identifying high-risk patients and optimal surgical parameters that, in the best-case scenario, could lead to further reduced rates of adverse events in the future.

AUTHOR CONTRIBUTIONS

All authors contributed to at least one of the following manuscript preparation roles: conceptualization AND/OR methodology, software, investigation, formal analysis, data curation, visualization, and validation AND drafting or reviewing/editing the final draft. As corresponding author, Dr Venäläinen confirms that all authors have provided the final approval of the version to be published and takes responsibility for the affirmations regarding article submission (eg, not under consideration by another journal), the integrity of the data presented, and the statements regarding compliance with institutional review board/Helsinki Declaration requirements.

REFERENCES

1. Peters RM, van Steenbergen LN, Stewart RE, et al. Patient characteristics influence revision rate of total hip arthroplasty: American Society of Anesthesiologists score and body mass index were the strongest predictors for short-term revision after primary total hip arthroplasty. J Arthroplasty 2020;35(1):188–192.e2.
2. Bozic KJ, Lau E, Ong K, et al. Risk factors for early revision after primary total hip arthroplasty in Medicare patients. Clin Orthop Relat Res 2014;472(2):449–454.
3. Jeschke E, Citak M, Günster C, et al. Obesity increases the risk of postoperative complications and revision rates following primary total hip arthroplasty: an analysis of 131,576 total hip arthroplasty cases. J Arthroplasty 2018;33(7):2287–2292.e1.
4. Rhee C, Lethbridge L, Richardson G, et al. Risk factors for infection, revision, death, blood transfusion and longer hospital stay 3 months and 1 year after primary total hip or knee arthroplasty. Can J Surg 2018;61(3):165–176.
5. Murphy BP d'S, Dowsey MM, Choong PFM. The impact of advanced age on the outcomes of primary total hip and knee arthroplasty for osteoarthritis: a systematic review. JBJS Rev 2018;6(2):e6.
6. Inacio MCS, Dillon MT, Miric A, et al. Mortality after total knee and total hip arthroplasty in a large integrated health care system. Perm J 2017;21(3):16–171.
7. Silman AJ, Combescure C, Ferguson RJ, et al. International variation in distribution of ASA class in patients undergoing total hip arthroplasty and its influence on mortality: data from an international consortium of arthroplasty registries. Acta Orthop 2021;92(3):304–310.
8. Ackerman IN, Bohensky MA, Zomer E, et al. The projected burden of primary total knee and hip replacement for osteoarthritis in Australia to the year 2030. BMC Musculoskelet Disord 2019;20(1):90.
9. Nemes S, Gordon M, Rogmark C, et al. Projections of total hip replacement in Sweden from 2013 to 2030. Acta Orthop 2014;85(3):238–243.
10. Sloan M, Premkumar A, Sheth NP. Projected volume of primary total joint arthroplasty in the U.S., 2014 to 2030. J Bone Joint Surg Am 2018;100(17):1455–1460.
11. Partridge T, Jameson S, Baker P, et al. Ten-year trends in medical complications following 540,623 primary total hip replacements from a national database. J Bone Joint Surg Am 2018;100(5):360–367.
12. Grosso MJ, Neuwirth AL, Boddapati V, et al. Decreasing length of hospital stay and postoperative complications after primary total hip arthroplasty: a decade analysis from 2006 to 2016. J Arthroplasty 2019;34(3):422–425.
13. Patel A, Pavlou G, Mújica-Mota RE, et al. The epidemiology of revision total knee and hip arthroplasty in England and Wales: a comparative analysis with projections for the United States. A study using the National Joint Registry dataset. Bone Joint J 2015;97-B(8):1076–1081.
14. Schwartz AM, Farley KX, Guild GN, et al. Projections and epidemiology of revision hip and knee arthroplasty in the United States to 2030. J Arthroplasty 2020;35(6S):S79–S85.
15. 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(2):574–583.
16. Kunutsor SK, Whitehouse MR, Blom AW, et al. Systematic review of risk prediction scores for surgical site infection or periprosthetic joint infection following joint arthroplasty. Epidemiol Infect 2017;145(9):1738–1749.
17. Paxton EW, Inacio MCS, Khatod M, et al. Risk calculators predict failures of knee and hip arthroplasties: findings from a large health maintenance organization. Clin Orthop Relat Res 2015;473(12):3965–3973.
18. Tan TL, Maltenfort MG, Chen AF, et al. Development and evaluation of a preoperative risk calculator for periprosthetic joint infection following total joint arthroplasty. J Bone Joint Surg Am 2018;100(9):777–785.
19. Venäläinen MS, Panula VJ, Klén R, et al. Preoperative risk prediction models for short-term revision and death after total hip arthroplasty: data from the Finnish Arthroplasty Register. JBJS Open Access 2021;6(1):e20.00091.
20. Bülow E, Hahn U, Andersen IT, et al. Prediction of early periprosthetic joint infection after total hip arthroplasty. Clin Epidemiol 2022;14:239–253.
21. Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. BMJ 2015;351:h3868.
22. van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014;14(1):137.
23. Havelin LI, Fenstad AM, Salomonsson R, et al. The Nordic Arthroplasty Register Association: a unique collaboration between 3 national hip arthroplasty registries with 280,201 THRs. Acta Orthop 2009;80(4):393–401.
24. Mäkelä KT, Matilainen M, Pulkkinen P, et al. Countrywise results of total hip replacement. An analysis of 438,733 hips based on the Nordic Arthroplasty Register Association database. Acta Orthop 2014;85(2):107–116.
25. Jakobsen JC, Gluud C, Wetterslev J, et al. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017;17(1):162.
26. Stavseth MR, Clausen T, Røislien J. How handling missing data may impact conclusions: a comparison of six different imputation methods for categorical questionnaire data. SAGE Open Med 2019;7:2050312118822912.
27. Belmont PJ Jr, Goodman GP, Hamilton W, et al. Morbidity and mortality in the thirty-day period following total hip arthroplasty: risk factors and incidence. J Arthroplasty 2014;29(10):2025–2030.
28. Hailer NP, Weiss RJ, Stark A, et al. The risk of revision due to dislocation after total hip arthroplasty depends on surgical approach, femoral

head size, sex, and primary diagnosis. An analysis of 78,098 operations in the Swedish Hip Arthroplasty Register. Acta Orthop 2012; 83(5):442–448.

29. Panula VJ, Alakylä KJ, Venäläinen MS, et al. Risk factors for prosthetic joint infections following total hip arthroplasty based on 33,337 hips in the Finnish Arthroplasty Register from 2014 to 2018. Acta Orthop 2021;92(6):665–672.

30. Panula VJ, Ekman EM, Venäläinen MS, et al. Posterior approach, fracture diagnosis, and American Society of Anesthesiology class III-IV are associated with increased risk of revision for dislocation after total hip arthroplasty: an analysis of 33,337 operations from the Finnish Arthroplasty Register. Scand J Surg 2021;110(3):351–358.

31. Thien TM, Chatziagorou G, Garellick G, et al. Periprosthetic femoral fracture within two years after total hip replacement: analysis of 437,629 operations in the Nordic Arthroplasty Register Association database. J Bone Joint Surg Am 2014;96(19):e167.

32. Arthursson AJ, Furnes O, Espehaug B, et al. Prosthesis survival after total hip arthroplasty—does surgical approach matter? Analysis of 19,304 Charnley and 6,002 Exeter primary total hip arthroplasties reported to the Norwegian Arthroplasty Register. Acta Orthop 2007; 78(6):719–729.

33. Lenguerrand E, Whitehouse MR, Beswick AD, et al. Risk factors associated with revision for prosthetic joint infection after hip replacement: a prospective observational cohort study. Lancet Infect Dis 2018; 18(9):1004–1014.

34. Lazarinis S, Mäkelä KT, Eskelinen A, et al. Does hydroxyapatite coating of uncemented cups improve long-term survival? An analysis of 28,605 primary total hip arthroplasty procedures from the Nordic Arthroplasty Register Association (NARA). Osteoarthritis Cartilage 2017;25(12):1980–1987.

35. Sarker IH. Machine learning: algorithms, real-world applications and research directions. SN Comput Sci 2021;2(3):160.

36. Venäläinen MS, Klén R, Mahmoudian M, et al. Easy-to-use tool for evaluating the elevated acute kidney injury risk against reduced cardiovascular disease risk during intensive blood pressure control. J Hypertens 2020;38(3):511–518.

37. Mahmoudian M, Venäläinen MS, Klén R, et al. Stable iterative variable selection. Bioinformatics 2021;37(24):4810–4817.

38. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-VerlagNew York; 2009.

39. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12(1):77.

40. Martin RK, Wastvedt S, Pareek A, et al. Predicting anterior cruciate ligament reconstruction revision: a machine learning analysis utilizing the Norwegian Knee Ligament Register. J Bone Joint Surg Am 2022; 104(2):145–153.

41. Dagneaux L, Amundson AW, Larson DR, et al. Contemporary mortality rate and outcomes in nonagenarians undergoing primary total hip arthroplasty. J Arthroplasty 2021;36(4):1373–1379.

42. Robinson J, Shin JI, Dowdell JE, et al. Impact of gender on 30-day complications after primary total joint arthroplasty. J Arthroplasty 2017;32(8):2370–2374.

43. Dale H, Børsheim S, Kristensen TB, et al. Perioperative, short-, and long-term mortality related to fixation in primary total hip arthroplasty: a study on 79,557 patients in the Norwegian Arthroplasty Register. Acta Orthop 2020;91(2):152–158.

44. Patsiogiannis N, Kanakaris NK, Giannoudis PV. Periprosthetic hip fractures: an update into their management and clinical outcomes. EFORT Open Rev 2021;6(1):75–92.

45. Franklin J, Malchau H. Risk factors for periprosthetic femoral fracture. Injury 2007;38(6):655–660.

46. Bloemheuvel EM, Van Steenbergen LN, Swierstra BA. Comparable mortality but higher revision rate after uncemented compared with cemented total hip arthroplasties in patients 80 years and older: report of 43,053 cases of the Dutch Arthroplasty Register. Acta Orthop 2022;93:151–157.

47. Thoen PS, Lygre SHL, Nordsletten L, et al. Risk factors for revision surgery due to dislocation within 1 year after 111,711 primary total hip arthroplasties from 2005 to 2019: a study from the Norwegian Arthroplasty Register. Acta Orthop 2022;93:593–601.

48. Madanat R, Laaksonen I, Graves SE, et al. Ceramic bearings for total hip arthroplasty are associated with a reduced risk of revision for infection. Hip Int 2018;28(3):222–226.

49. Mäkelä KT, Furnes O, Hallan G, et al. The benefits of collaboration: the Nordic Arthroplasty Register Association. EFORT Open Rev 2019; 4(6):391–400.

50. Smith JO, Frampton CMA, Hooper GJ, et al. The impact of patient and surgical factors on the rate of postoperative infection after total hip arthroplasty-a New Zealand Joint Registry Study. J Arthroplasty 2018;33(6):1884–1890.

51. Trela-Larsen L, Kroken G, Bartz-Johannessen C, et al. Personalized estimation of one-year mortality risk after elective hip or knee arthroplasty for osteoarthritis. Bone Joint Res 2020;9(11):808–820.

52. Garland A, Bülow E, Lenguerrand E, et al. Prediction of 90-day mortality after total hip arthroplasty. Bone Joint J 2021;103-B(3): 469–478.