# Preoperative prediction of medical morbidity after fast-track hip and knee arthroplasty, a machine learning based approach.

**2. Author information:**

Christian Michelsen, M.Sci., Research Fellow, The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17 2100 Copenhagen, Denmark

Christoffer C Jørgensen, M.D., Senior Researcher, Section of Surgical Pathophysiology and Centre for Fast-track Hip and Knee Replacement, 7621, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen, Denmark

Mathias Heltberg, M.Sci, Research Fellow, The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17 2100 Copenhagen, Denmark

Mogens H Jensen, D.Sci., Prof., The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17 2100 Copenhagen, Denmark

Alessandra Luchetti, M.Sci., Research Fellow The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17 2100 Copenhagen, Denmark

Pelle B Petersen2, M.D., Ph.D, Section of Surgical Pathophysiology and Centre for Fast-track Hip and Knee Replacement, 7621, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen, Denmark

Troels Petersen, M.Sci, Ass.Prof., The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17 2100 Copenhagen, Denmark

Henrik Kehlet,2 M.D., DM.Sci., Prof. Section of Surgical Pathophysiology and Centre for Fast-track Hip and Knee Replacement, 7621, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen, Denmark

\*This is a joint first-authorship between CM and CJ

**3. Corresponding author:**

Dr. Christoffer Calov Jørgensen

Section for Surgical Pathophysiology 7621

Rigshospitalet, Blegdamsvej 9,

DK-2100 Copenhagen, Denmark

Phone +45 3545 4616 Fax: +45 3545 6543

E-mail: christoffer.calov.joergensen@regionh.dk

**4. Clinical Trial Number:** The Centre for Fast-track Hip and Knee Replacement Database was registered as a study registry on ClinicalTrials,gov:NCT01515670

**5. Prior presentations:** Not applicable

**6. Acknowledgements**: The members of the Centre for Fast-track Hip and Knee Replacement Database collaborative group all contributed by implementing the fast-track protocol at their respective departments and reviewing the final manuscript.

Frank Madsen M.D. Consultant, Department of Orthopedics, Aarhus University Hospital, Aarhus, Denmark

Torben B. Hansen M.D., Ph.D., Prof. Department of Orthopedics, Holstebro Hospital, Holstebro, Denmark

Kirill Gromov, M.D., Ph.D., Ass.Prof. Department of Orthopedics Hvidovre Hospital, Hvidovre Denmark

Thomas Jakobsen, M.D., Ph.D., DM.Sci., Ass. Prof. Department of Orthopedics, Aalborg University Hospital, Farsø, Denmark

Claus Varnum, M.D., Ph.D., Ass. Prof. Department of Orthopedics, University of Southern Denmark Vejle, Denmark

Soren Overgaard, M.D., DM.Sci., Prof, Department of Orthopedics, Bispebjerg Hospital, Copenhagen, Denmark

Mikkel Rathsach, M.D., Ph.D., Ass. Prof. Department of Orthopedics, Gentofte Hospital, Gentofte, Denmark

Lars Hansen, M.D., Consultant, Department of Orthopedics, Sydvestjydsk Hospital, Grindsted, Denmark

**7. Word and Element Counts:**

Abstract: 299/300 Introduction: 408/500 Discussion:1236/1500 Figures:3 Tables:2 Appendices:1 Supplementary Digital Files:4

**8. Abbreviated title:** Machine learning for prediction of morbidity in THA/TKA

**9. Summary Statement:**  Not applicable.

**10. Funding**: The study received funding from the Lundbeck Foundation, Denmark, as well as from institutional and departmental sources.

**11. Conflict of interest**: Prof. Kehlet is a board member of “Rapid Recovery”, by Zimmer Biomet. Mr. Heltberg is sponsored by a grant from the Lundbeck Foundation, independently of the present study.

# Abstract

**Background**: There are many efforts to predict postoperative LOS and morbidity. Recently, machine-learning (ML) methods have potentially improved surgical risk prediction, including after total hip (THA) and knee arthroplasty (TKA). However, few studies include enhanced recovery programs, and most rely on administrative coding with limited follow-up and information on perioperative care. Thus, benefits of ML-methods for prediction of postoperative morbidity in enhanced recovery THA and TKA are uncertain.

**Methods:** Multicenter cohort study in enhanced recovery THA and TKA. Prospective recording of comorbidity and prescriptions. Information on LOS and readmissions through the Danish National Patient registry and medical records. Data was split into a training (n:18013 surgery 2014-2016) and a test set (n:3913 surgery in 2017). A LightGBM ML-method using 33 variables (ML33) was used for predicting medical morbidity with a LOS >4 days or 90-days readmission and compared to a logistic regression (LR33) model, a ML-model excluding age, an Age-model and two models (ML10 and LR10) using only the top ten variables. Model performances were evaluated using various metrics, including positive predictive value (PPV), operating receiver (AUROC) and precision recall curves (AUPRC). Variable importance was analyzed using SHAP values.

**Results:** With 782 “risk-patients”, AUROC and AUPRC were 76.3 and 75.9 and 15.5 and 17.1 for the ML33 and ML10, respectively, vs. 74.5 and 15.7 for LR33. PPV were 13.6% and 12.8% for ML33 and ML10 vs. 12.5% for LR33. The ML-model excluding age and Age-model performed worst. SHAP analyses found overlapping variable importance between ML33 and LR33, and with considerable variations in influence of prescribed medicine

**Conclusion:** ML-algorithms may improve prediction of patients in high-risk of medical complications in fast-track THA and TKA compared to a logistic regression model. Future studies could benefit from such algorithms to find a manageable population of patients benefitting most from intensified perioperative care.

# INTRODUCTION

Prediction of postoperative morbidity and requirement for hospitalization is important for planning of health care resources. With regard to the common surgical procedures of primary total hip and knee arthroplasty (THA and TKA), the introduction of enhanced recovery or fast-track programs have led to a significant reduction of postoperative length of stay (LOS) as well as morbidity and mortality.1-4 However, despite such progress, a fraction of patients still have postoperative complications leading to prolonged LOS or readmissions.1,4,5 Consequently, in order to prioritize perioperative care, many efforts have been published to preoperatively predict LOS and morbidity using traditional risk factors such as age, preoperative cardio-pulmonary disease, anemia, diabetes, frailty, etc.5-10 These efforts have been based on traditional statistical methods, most often multiple regression analyses, and essentially concluding that it is “better to be young and healthy than old and sick”. Consequently, despite being statistically significant, conventional risk-stratification based on such studies have had a relatively limited clinically relevant ability to predict potentially preventable morbidity and LOS.5-11

More recently, machine learning (ML) methods have been introduced with success in several areas of healthcare and where preliminary data suggest them to improve surgical risk prediction compared to traditional risk calculation in certain anesthetic and surgical conditions12,13 This is also the case in both THA, TKA and uni-compartmental knee replacement (UKA), where several publications on ML-algorithms for prediction of LOS14,15, complications,16 disability,17 potential outpatient setup,18 readmissions19 or payment models20,21, have shown promising predictive value compared to conventional statistical methods.22

However, few of such papers have included enhanced recovery programs, and most are based on large database cohorts with the presence of risk factors and complications often relying on administrative coding with limited information on perioperative care, follow-up and discharge destination. Thus, a recent study THA and TKA within an enhanced recovery protocol and including the above information, did not find advantages of ML methods compared to logistic regression in predicting a LOS > 2 days.23 Nonetheless, ML predictive models remain promising and could provide an improved basis for identifying a potential “high-risk” surgical population who may benefit from more extensive preoperative evaluation and postoperative medical care.

Consequently, we analyzed whether a ML-algorithm was able to improve preoperative prediction of medical complications resulting in prolonged LOS and readmissions compared to a traditional logistic regression model in a large, consecutive cohort of patients having fast-track THA/TKA in a national public health-care system.1 In addition to well-defined patient-reported preoperative risk-factors, we included information on reimbursed prescriptions 6 months prior to surgery using a nationwide registry.24

# Method

Reporting of the study is done in accordance with the Transparent reporting of multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement25 and the Clinical AI Research (CAIR) checklist proposal.26  
The study is based on the Centre for Fast-track Hip and Knee Replacement database which is a prospective database on preoperative patient characteristics and enrolling consecutive patients from 7 departments between 2010 and 2017. The database is registered on ClinicalTrials.gov as a study registry (NCT01515670). Permission to review and store information from medical records without informed consent was acquired from Center for Regional Development (R-20073405) and the Danish Data Protection Agency (RH-2007-30-0623). Patients completed the preoperative questionnaire with nurse assistance if needed and additional information on reimbursed prescriptions 6 months prior to surgery was acquired using the Danish National Database of Reimbursed Prescriptions (DNDRP) which records all dispensed prescriptions with reimbursement in Denmark.24 Finally, data was crossed with the Danish National Patient Registry (DNPR) for information on LOS (counted as nights spent in hospital), 90-days readmissions with overnight stay and mortality. In case of LOS >4 days or readmission, patient discharge summaries were reviewed for information on postoperative morbidity and in case of insufficient information, the entire medical records were reviewed. Readmissions were only included if considered related to the surgical procedure, thus excluding planned procedures like cancer workouts, cataract surgery, etc. Readmissions due to urinary tract infection or dizziness after day 30 were also considered unrelated to the surgical procedure. In case of postoperative mortality, the entire medical record, including potential readmissions was always reviewed to identify cause of death. Evaluation of discharge and medical records was performed by CJ until 2015, and then by PP supervised by CJ. In case of disagreement, records were conferred with HK. Subsequently, causes of LOS >4, readmissions or mortality were classified as “medical” when related to perioperative care (renal failure, falls, pain, thrombosis, anemia, venous thromboembolism or infection etc.) and “surgical” if related to surgical technique (prosthetic infection, revision surgery, periprosthetic fracture, hip dislocation etc.).1 In case of a LOS >4 days with a standard discharge summary describing a successful postoperative course, it was assumed that no clinically relevant postoperative complications had occurred.   
For the present study, only cases between 2014 and 2017 were used to provide the most up-to date data. All patients had elective unilateral THA or TKA in dedicated arthroplasty departments with similar fast-track protocols, including multimodal opioid sparing analgesia with high-dose (125mg) methylprednisolone, preference for spinal anesthesia, only in-hospital thromboprophylaxis when LOS ≤5 days, early mobilization, functional discharge criteria and discharge to own home.1 There are no selection criteria for the fast-track protocol as it is considered standard of care, but we excluded patients with previous major hip or knee surgery within 90-days of their THA or TKA, THA due to severe congenital joint disorder and cancer surgery.

## Outcomes

The primary outcome was to compare prediction quality when using an ML algorithm to predict the occurrence of “medical” complications resulting in a LOS >4 days or readmission compared to a traditional logistic regression model (outcome A). Secondarily, we investigated how inclusion of cases with a LOS >4 days but no reported medical complication as positive outcome influenced the model (outcome B). For both outcomes, we also investigated whether a parsimonious model with a limited number of covariates would equally performant to the full-scale model with all available variables and whether the effect of age per se would compare to the full-scale ML-model. All figures and tables in the main text are based on outcome A; the corresponding figures for outcome B can be found in the Supplemental Digital Content.

## Statistical Analysis

Data was initially trimmed by removing 156 patients (1.7%) who either had missing data or were outliers with regards to weight (<30 kg or >250 kg) and height (<100 cm or >210 cm). To reduce the risk of overfitting, data was split into a training set consisting of 18.013 procedures with surgery between 2014-2016 and a test set of 3913 procedures with surgery in 2017.

As a reference model, we used classical logistic regression (LR) using all 33 input variables (LR33). Cases of missing values in the LR were handled by imputing missing values with the median of present values. All variables were then normalized.

In addition, we used Boosted Decision Trees (LightGBM)27 for the ML models as such methods work well with categorical data and missing values, unlike neural networks. We tried using both normal cross entropy and FocalLoss28 as the objective function for the ML model. The reason behind also testing FocalLoss was to allow the ML model to focus more on the (few) positives.

The ML model based on all input variables (ML33) was trained and hyperparameter optimized using state of the art ML methods. The models were trained on the training data and then used for making predictions on the unseen test data (see supplementary for details). The classification threshold was calibrated such that no more than 20% of the total number of patients are predicted by the model to experience complications, i.e. a positive predictive fraction (PPF) of 20%. We also included results for PPF values of 25% and 30%. We also trained two parsimonious models using ML and LR with only the 10 most important features (ML10, LR10). Finally, we specifically explored the influence of increasing age, by constructing a model based only on age (Age), and a ML model based on all variables but age (ML-NoAge).

To investigate the importance of the included variables, we computed the SHapley Additive exPlanations (SHAP) values, which provide estimates on which variables contribute most to the risk score predictions.29,30 We also investigated potential relation between reimbursed prescribed cardiac drugs, anticoagulants, psychotropics and pulmonary drugs and age, as investigations on relation between polypharmacy and postoperative outcomes has mainly been in older patients.31  
For evaluating model performance, we computed the number of true positives (TP), false positives (FP), false negatives (FN), true negatives (TN), true positive rate (TPR = sensitivity), positive predictive value (PPV = precision). Since the data is quite imbalanced (about a 1:20 positive:negative ratio) we also computed the Matthews Correlation Coefficient (MCC) which is independent of class imbalance.32,33 The MCC ranges between -1 (the 100% wrong classifier), 0 (the random classifier), and +1 (the perfect classifier). Finally, we computed the area under the receiver operating characteristic curve (AUROC) and the area under the precision recall curve (AUPRC). To evaluate the statistical difference between the classifiers, we applied a Bayesian metric comparison.34 This calculates P(TPR) , which is the probability that a model will perform better than the ML model relative to the TPR metric. For two equal models, the P(TPR) value would be approximately 50%.

# **Results**

Of the 3913 patients median age was 70 years (IQR 62-76), 59% were female and 58% had THA (table 1) Details on prescribed drug types can be found in Appendix 1. Median LOS was 2 (IQR: 1-2) days with 7.6% 90-days readmissions and outcome A occurring in 182 (4.7%) patients. When applying any model with a PPF of 20% to the 3913 patients, 782 qualified as “risk-patients”. The results are summarized in figure 1 and table 2. When considering risk scores from ML33 (figure 1a) and LR33 leading to this risk-patient selection, 106 and 98 had outcome A (TP), respectively. Correspondingly, the TPR and PPV was 58.2% and 13.6% for ML33 and 53.8% and 12.5% for LR33. ML33 was superior (figure 1b) on essentially all parameters compared to any of the other models, although the differences were minor (table 2).The results were similar when using a PPF of 25% and 30%, but with the TPR for ML33 increasing to 64.4% and 69.2% and PPV decreasing to 12.0% and 10.7%, respectively.

Both the ML-NoAge and Age models had significantly lower TPR than the ML33 model (figure 1b). Despite age being the single most important variable (figure 2), the ML model excluding age (ML-NoAge) performed as well as age alone.

When evaluating feature importance, we found a strong correlation between ML33 and LR33 with age and use of walking aids being the most important variables in both models (figure 2a). From the combined importance of variables outside the top ten, it seemed that the ML approach extracted more information with fewer variables than the LR approach (figure 1b).

In ML33 specifically, there was a clear signal that increasing age, number of reimbursed prescriptions, and presence of comorbidity all contributed to an increased risk score. In contrast, recent date of surgery and increased hemoglobin level seemed to reduce the calculated risk (figure 2b).

Individual analysis of the SHAP interaction values for types of anticoagulant prescriptions revealed that prescriptions on VKA or ADP-antagonists increased, and acetylic salicylic acid (ASA) and direct oral anticoagulants (DOAC) reduced the ML33 risk score, regardless of age (figure 3a). SHAP analysis of prescribed cardiac drugs revealed that prescriptions on Calcium antagonists and betablockers in combination with other antihypertensives increased ML33 risk, as did prescriptions on nitrates, other antihypertensives and antiarrhythmics. For the remaining cardiac drugs, prescriptions either reduced or had minor influence on the ML33 risk score, and there was hardly any relation with age (figure 3b). Having psychotropic prescriptions increased the ML33 risk score except for those using antipsychotics (0.6%) which had no influence on the risk score. For users of selective serotonin inhibitors there was a clear age-related distinction with the ML33 risk score being increased in the elderly patients but decreased in those < 60 years (figure 3c). Finally, ML33 risk score was increased in those with prescriptions on inhalation steroid and beta-agonists, however there was a clear association with increased age as the risk increase was most expressive in the younger patients (figure 3d). The results including patients with a LOS >4 days, but no reported postoperative complications (outcome B) were similar as for outcome A. In general, we found that ML33 was superior to all the other models, however, with a lower performance increase than in outcome A. (Supplemental Digital Context table S1 and Figure S2a+b) However, the SHAP analysis on prescriptions found much greater variations across different types of drugs. Thus, the reduced risk with ASA and DOAC prescriptions was attenuated, as was the influence of practically all cardiac drugs except for nitrates, other antihypertensives and antiarrhythmics (Supplemental Digital Context figure S3a-d).

# Discussion

We found that using a ML algorithm including all 33 available variables and a parsimonious ML-algorithm encompassing only the 10 most important predictors improved prediction of patients at increased risk of having a LOS >4 days or readmissions due to medical complications compared to corresponding traditional logistic regression models. In contrast, when also including patients having a LOS >4 days, but without a well-defined complication as an outcome, the ML33 remained superior, but the parsimonious ML-algorithm was slightly worse than a traditional logistic regression model including all variables (LR33). We also found that although age was the single most important predictor of both outcome A and B, it was not well suited for prediction of postoperative medical complications after fast-track THA and TKA on its own. Finally, we demonstrated how the chosen classification threshold (PPF) of the ML algorithm influenced model performance through an increase in TPR at the cost of decreased PPV.

A previous systematic review also found that ML algorithms can provide accurate prediction of postoperative outcomes in THA and TKA.35 However, the authors concluded that such models performed best at predicting postoperative complications, pain and patient reported outcomes but were less accurate at predicting readmissions and reoperations.35 That ML algorithms may improve prediction of complications after THA and TKA compared to traditional logistic regression was also found by Shah et al. who used an automated ML framework to predict selected major complications after THA.16 However, it was a retrospective study based on diagnostic and administrative coding and the selected complications occurred only in 0.61% of patients, potentially limiting clinical relevance. In contrast, we aimed at identifying a cohort which would comprise 20% of patients in which we then find about 60% of all medical complications. This we believe is within the means of the Danish socialized healthcare system to allocate additional resources for intensified perioperative care and with both patient-related and economic benefits due to potentially avoided complications and costs.  
In contrast to many other ML studies,36 our dataset included only preoperative data and the only paraclinical data was preoperative hemoglobin. Although the inclusion of other laboratory tests such as preoperative albumin, sodium and alkaline phosphatase has been found to be of importance in ML algorithms for home discharge in UKA15 and spine surgery,12 they are not standard in our fast-track protocols. Furthermore, as there is a need to prioritize the limited health-care resources, most decisions on which patients may benefit from more extensive postoperative care will likely need to be conducted preoperatively. Thus, although postoperative information such as duration of surgery, perioperative blood loss or postoperative hemoglobin have been included in other studies36, we decided against the use of any peri- and postoperative data. The same approach has been used by Ramkumar and colleagues who used U.S. National Inpatient Sample data including 15 preoperative variables, to predict LOS, patient charges and disposition after both TKA37 and THA.21 However, these studies were not conducted in a socialized health care system, and the main focus was on the need for differentiated payment bundles and without specific information on reason for increased LOS or non-home discharge.37 Wei et al. used an artificial neural network model to predict same-day discharge after TKA, based on the NSQUIP database from 2018. They also compared with logistic regression and found that six of the ten most important variables were the same, similar to our findings.38 However, patients with one-day LOS were intentionally excluded due to variations in in-patient vs. out-patient registration.38   
Age has traditionally been a major consideration with regards to surgical procedures which is why we choose to specifically evaluate its effect on our risk-prediction. That age is important for risk-prediction was further illustrated by ML-noAge being comparable to the Age model. Note that, although elderly patients had increased risk of postoperative complications, likely related to decline of physical reserves,39 using chronological age alone as a selection criteria for being a surgical “risk-patient” was much inferior compared to both ML and LR models incorporating comorbidity and functional status. We used the SHAP values for estimation of the feature importance, thus providing a better understanding of the otherwise “black-box” ML model. SHAP values show which variables contribute most to the risk score predictions. In this context our inclusion of specific data on reimbursed prescriptions 6 months prior to surgery unsurprisingly found increased risk-scores with increased number of prescriptions and that most prescriptions were in elderly patients. A Canadian study in elective non-cardiac surgery found decreased survival and increased LOS and readmissions and costs in patients >65 years with polypharmacy.31 However, this is a complex relationship where some patients benefit from their treatment while other may suffer from undesirable side-effects and the authors cautioned against altering perioperative practices based on current evidence.31 However, the information from the included prescriptions with SHAP analysis may provide inspiration for new hypothesis-generating studies investigating i.e. cause of the age related differences in risk score for SRRIs or the potential differences in risk-profile between having preoperative prescribed VKA and DOAK as seen in our study.

Another requirement for ML-algorithms to be clinically useful is user friendliness and not depending on excessive additional data collection by the attending clinicians. In this context, it was a bit disappointing that the parsimonious ML algorithm with only the 10 most important variables was slightly worse at predicting outcome B than LR33. A reason for this could be that when including a LOS >4 days but without described medical complications, the combination of all variables provides information not available by merely including the ten most important ones. This highlights the need for as much detailed data as possible in order to fulfill the true potential of ML algorithms.

Our study has some limitations. We included only a limited number of, often binary, preoperative variables. As analysis of multilevel continuous data is one of the strengths of ML compared to logistic regression, this may limit full realization of our ML model. As previously mentioned, we also excluded intraoperative information, including type of anesthesia, prosthesis etc. all of which may influence postoperative outcomes. The observational design of this study means that we cannot exclude unmeasured confounding or confounding by indication. Also, despite that the DNDRP has a near complete registration of dispensed medicine in Denmark, some types or drugs, especially benzodiazepines, are exempt from general reimbursement and thus not sufficiently captured.24 Also, it is doubtful that patient used all types of drugs at time of surgery (e.i. heparin which is rarely for long-term use). Finally, classification of a complication being “medical” depended on review of the discharge records which can also introduce bias. However, we believe this approach to be superior to depending on diagnostic codes which often are inaccurate and provide limited details on whether the complication may be attributed to a medical or surgical adverse event. The strengths of our study include use of national registries with high degree of completion (>99% of all somatic admissions in case of the DNDRP),40 prospective recording of comorbidity, extensive information on prescription patterns 6 months prior to surgery and similar established enhanced recovery protocols in all departments.

In summary, our results indicate that ML-algorithms may provide clinically relevant improved predictions for defining patients in high-risk of medical complications after fast-track THA and TKA compared to a logistic regression model. Future studies could benefit from using such algorithms to find a manageable population of patients who may benefit the most from intensified perioperative care.

# Tables

|  |  |  |
| --- | --- | --- |
| Table 1. patient demographics with and without outcome A in the combined test and training dataset. | | |
| Preoperative characteristics n(%) unless otherwise specified | +outcome A (n:1180) | -outcome A (n:20837) |
| mean age (SD) | 75.0 (68.0-81.0) | 69.0 (62.0-75.0) |
| mean number of reimbursed prescriptions1 (SD) | 3.0 (1.0-4.0) | 2 (0.0-3.0) |
| female gender | 755 (64.0) | 12133 (58.2) |
| THA | 636 (53.9) | 11542 (55.4) |
| mean weight in kg (SD) | 78.0 (67.0-91.0) | 81 (70.0-93.0) |
| mean height in cm (SD) | 168 (162.0-175.0) | 170.0 (164.0-178.0) |
| mean body mass index (SD) | 27.3 (23.9-31.2) | 27.5 (24.6-31.1) |
| regular use of walking aid  missing | 552 (46.8)  29 (2.5) | 4398 (21.5)  359 (1.7) |
| living alone  with others  institution  missing | 578 (49.0)  571 (48.4)  24 (2.0)  7 (0.6) | 6717 (32.2)  13869 (66.6)  113 (0.5)  138 (0.7) |
| hemoglobin  missing | 8.2 (7.7-8.8)  11 (0.9) | 8.6 (8.1-9.2)  314 (1.5) |
| >2 units of alcohol/day  missing | 79 (6.7)  10 (0.8) | 1589 (7.6)  174 (0.8) |
| active smoker  missing | 130 (11.0)  11 (0.9) | 2751 (13.2)  141 (0.7) |
| cardiac disease  missing | 306 (25.9)  8 (0.8) | 2750 (13.2)  153 (0.7) |
| hypercholesterolemia  missing | 467 (39.6)  8 (0.7) | 6062 (29.1)  120 (0.6) |
| hypertension  missing | 738 (62.5)  64 (5.4) | 10141 (48.7)  663 (3.2) |
| pulmonary disease  missing | 182 (15.4)  5 (0.4) | 1841 (8.8)  96 (0.5) |
| previous cerebral attack  missing | 165 (14.0)  25 (2.1) | 1086 (5.2)  282 (1.4) |
| previous VTE  missing | 133 (11.3)  26 (2.2) | 1481 (7.1)  325 (1.6) |
| malignancy (undefined)  previous radically treated malignancy  missing | 557 (47.2)  127 (10.8)  14 (1.2) | 8843 (42.4)  2065 (9.9)  162 (0.8) |
| chronic kidney disease  missing | 50 (4.2)  35 (3.0) | 273 (1.3)  292 (1.4) |
| family member with VTE  missing | 155 (13.1)  1190 (16.1) | 2510 (12.0)  2569 (12.3) |
| regular snoring  uncertain about snoring  missing | 266 (22.5)  208 (17.6)  259 (21.9) | 5522 (26.5)  3781 (18.1)  3309 (15.9) |
| not feeling rested  uncertain about being rested  missing | 468 (39.7)  48 (4.1)  105 (8.9) | 9340 (44.8)  809 (3.9)  1230 (5.9) |
| psychiatric disorder  missing | 156 (13.2)  62 (5.3) | 1590 (7.6)  703 (3.4) |
| Characteristic based on combination of questionnaire and DNDRP | | |
| diabetes  diet treated diabetes2  oral antidiabetics  insulin treated diabetes3  missing | 29 (2.5)  137 (11.6)  60 (5.1)  7 (0.6) | 274 (1.3)  1448 (6.9)  413 (2.0)  98 (0.5) |
| SD: standard deviation THA: total hip arthroplasty VTE: venous thromboembolic event DNDRP: Danish National Database of Reimbursed Prescriptions.  1Antirheumatica, steroids, anticoagulants, cardiac, cholesterol lowering, respiratory and psychotropic drugs. 2Reported diabetes but no registered prescriptions 2 +/- oral antidiabetics | | |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Performance of the six different models with positive prediction fractions of 20%, 25% and 30% | | | | | | | | | | |
| PPF 20 | TP | FP | FN | TN | TPR | PPV | MCC | AUROC | AUPRC | P (TPR) |
| ML33 | 106 | 676 | 76 | 3055 | 58.2% | 13.6% | 21.1% | 76.3% | 15.5% | - |
| LR33 | 98 | 684 | 84 | 3047 | 53.8% | 12.5% | 18.7% | 74.5% | 15.7% | 19.7% |
| ML10 | 100 | 682 | 82 | 3049 | 54.9% | 12.8% | 19.3% | 75.9% | 17.3% | 26.1% |
| LR10 | 95 | 687 | 87 | 3045 | 52.2% | 12.1% | 17.8% | 73.7% | 13.6% | 12.4% |
| ML-NoAge | 88 | 694 | 94 | 3037 | 48.4% | 11.3% | 15.7% | 72.3% | 13.6% | 3.1% |
| Age\_only | 87 | 676 | 95 | 3055 | 47.8% | 11.4% | 15.8% | 69.7% | 12.1% | 2.3% |
| PPF 25 |  |  |  |  |  |  |  |  |  |  |
| ML33 | 117 | 861 | 65 | 2870 | 64.3% | 12.0% | 20.0% | 76.3% | 15.5% | - |
| LR33 | 110 | 868 | 72 | 2863 | 60.4% | 11.2% | 18.1% | 74.5% | 15.7% | 23.1% |
| ML10 | 115 | 863 | 67 | 2868 | 63.2% | 11.8% | 19.5% | 75.9% | 17.3% | 41.2% |
| LR10 | 106 | 872 | 76 | 2859 | 58.2% | 10.8% | 17.0% | 73.4% | 15.5% | 11.8% |
| ML-NoAge | 106 | 872 | 76 | 2859 | 58.2% | 10.8% | 17.0% | 72.3% | 13.6% | 11.8% |
| Age\_only | 94 | 824 | 88 | 2907 | 51.6% | 10.2% | 14.7% | 69.7% | 12.2% | 0.7% |
| PPF 30 |  |  |  |  |  |  |  |  |  |  |
| ML33 | 126 | 1047 | 56 | 2684 | 69.2% | 10.7% | 18.9% | 76.3% | 15.5% | - |
| LR33 | 120 | 1053 | 62 | 2678 | 65.9% | 10.2% | 17.3% | 74.5% | 15.7% | 25.2% |
| ML10 | 124 | 1049 | 58 | 2682 | 68.1% | 10.6% | 18.4% | 75.9% | 17.3% | 40.8% |
| LR10 | 115 | 1058 | 67 | 2673 | 63.2% | 9.8% | 16.0% | 73.7% | 15.5% | 11.1% |
| ML-NoAge | 116 | 1057 | 66 | 2674 | 63.7% | 9.9% | 16.3% | 72.3% | 13.6% | 13.8% |
| Age\_only | 100 | 955 | 82 | 2776 | 54.9% | 9.5% | 13.9% | 69.7% | 12.2% | 0.2% |
| PPF: positive prediction fraction TP: true positives FP: false positives FN: false negatives TN: true negatives TPR: true positive rate PPV: positive predictive value MCC: Matthews correlation coefficient AURC: area under the ROC curve AUPRC: area under the precision recall curve P(TPR): probability that the model performs better than the ML model relative to TPR. Green/red colors indicates the model with the best/worst performance given that specific metric | | | | | | | | | | |

# Figures

Figure 1a+b

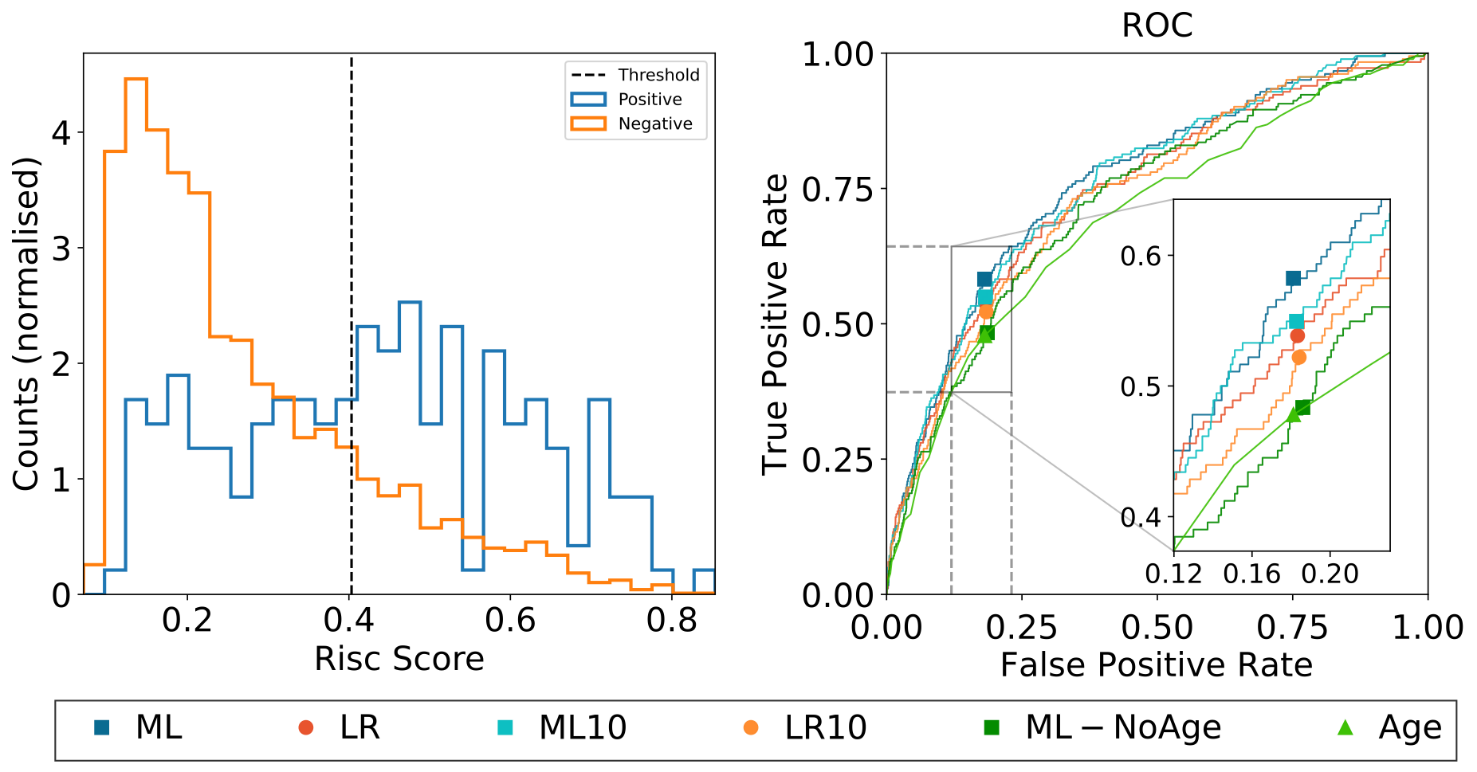
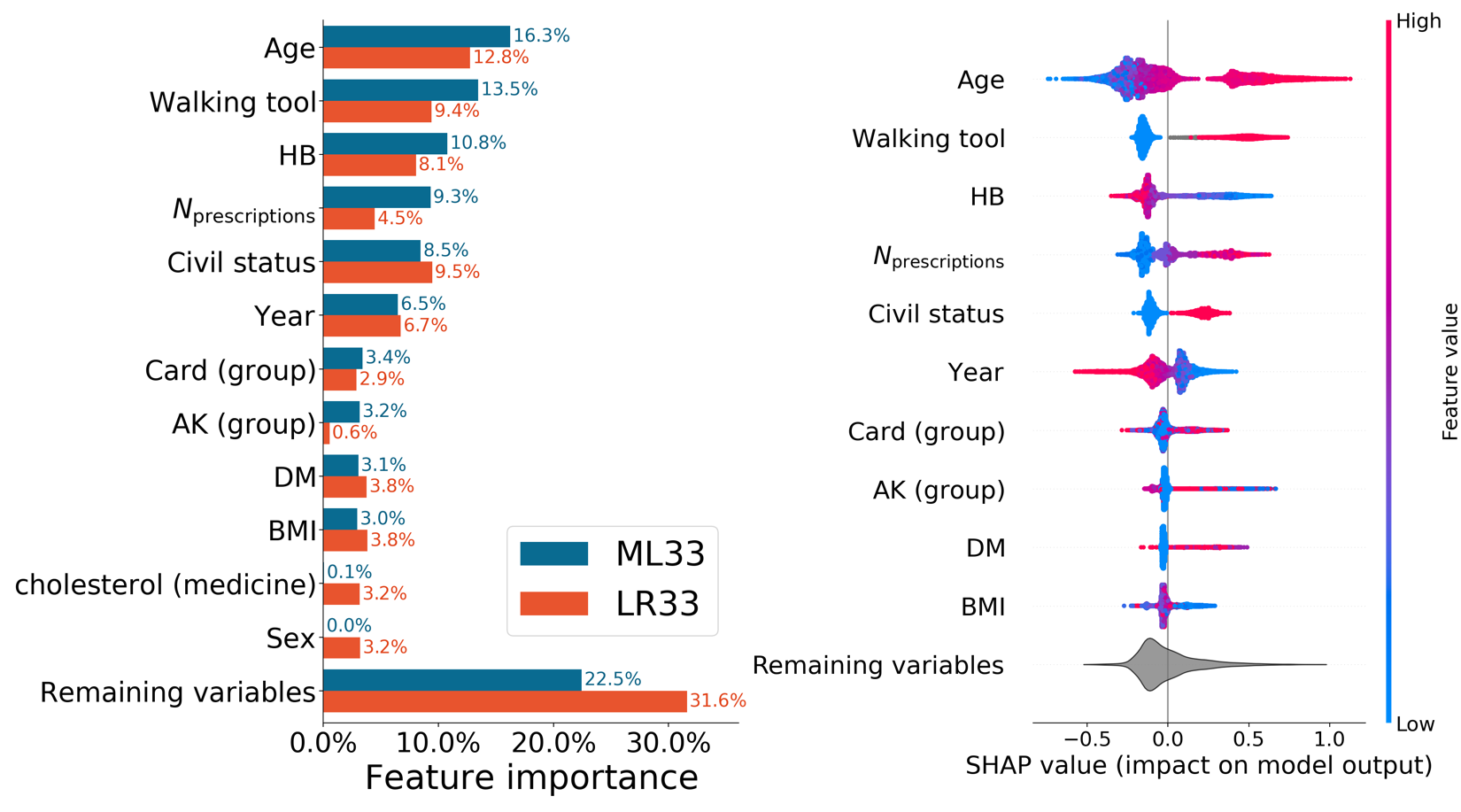


Figure 1a) Distribution of risk scores for patients +/- outcome A. The dashed line shows the classification threshold of 20% positive prediction fraction. Figure 1b) ROC curves for the complete machine learning model (ML33), complete logistic regression model (LR33), parsimonious machine learning model (ML10), parsimonious logistic regression model (LR10), machine learning using without age (ML-NoAge) and the age only model (Age).

Figure 2a+b

2a) The overall importance of the 10 most important variables measured by the SHAP values for both the ML33 and LR33 models. The contributions of the remaining variables are summed in the bottom bar. 2b) The SHAP values for the ML33 model. Positive SHAP-values increase the ML33 risk score while negative values decrease ML33 risk score. The color is related to the value of the variable with blue being lowest and red highest.

Figure 3 a+b



SHAP beeswarm plot on the contributions to the ML33 model of individual types of prescribed drugs stratified by age.

3a) Anticoagulants 3b) Cardiac drugs 3c) Psychotropics 3d) Respiratory drugs (NB stofgrupperne kommer på snarest)

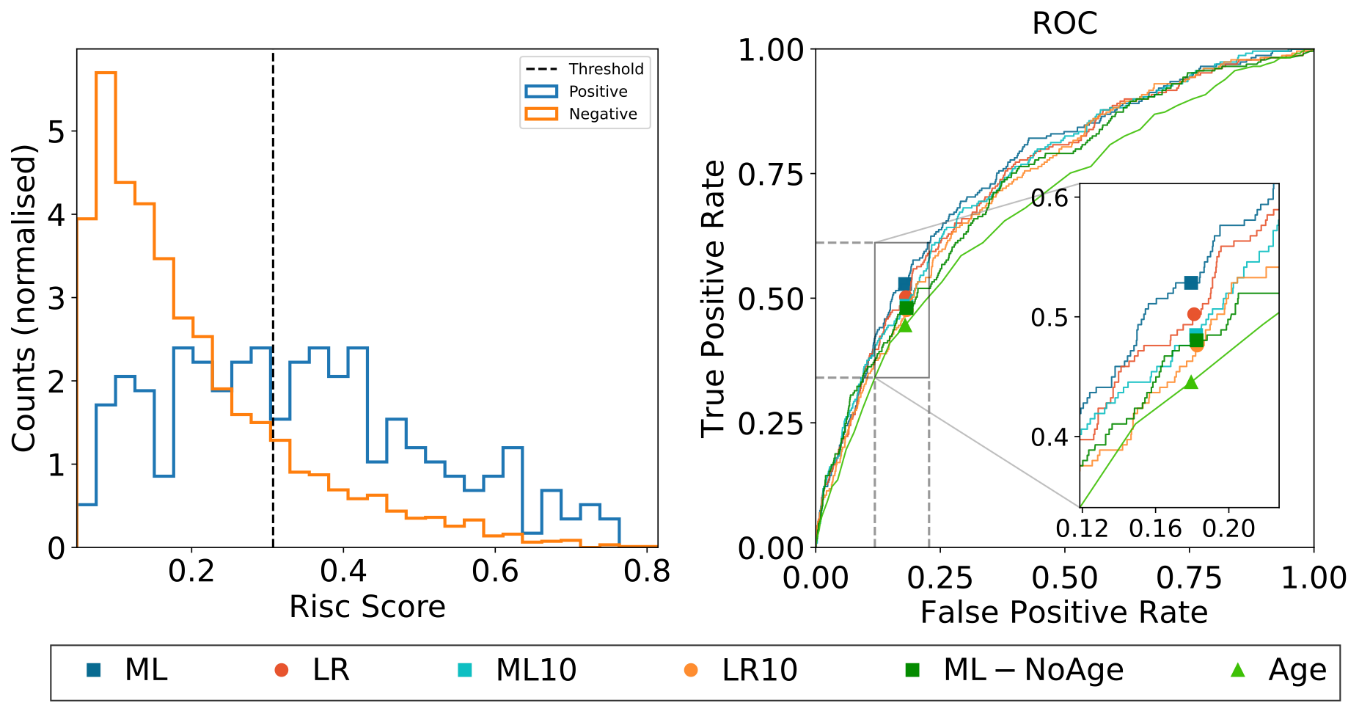
# Appendix

|  |  |  |
| --- | --- | --- |
| Appendix table 1. Details on specific drugs with reimbursed prescriptions 6 months preoperatively.  Numbers are n (%) | | |
| Anticoagulants  none  VKA  Heparin+ASA  DOAC  ASA  Dipyradimol  ADP  ASA+Dipyradimol  VKA+ASA  DOAC+ASA  VKA+ADP  DOAC+ADP  VKA+Heparin  DOAC+ASA+ADP  ASA+ADP  ASA+ADP+Heparin  ASA+ADP+Dipyradimol | 679 (57.5)  106 (9.0)  0 (0.0)  48 (4.1)  205 (17.4)  5 (0.4)  75 (6.4)  17 (1.4)  10 (0.8)  6 (0.5)  4 (0.3)  3 (0.3)  1 (0.1) 1 (0.1)  18 (1.5)  1 (0.1)  1 (0.1) | 15844 (76.0)  750 (3.6)  7 (0.0)  659 (3.2)  2492 (12.0)  29 (0.1)  569 (2.7)  168 (0.8)  78 (0.4)  41 (0.2)  11 (0.1)  14 (0.1)  21 (0.1)  3 (0.0)  132 (0.6)  12 (0.1)  7 (0.0) |
| Cardiac drugs  none  diuretics  angiotensin-II/ACE-inhibitors  calcium antagonists  betablockers  nitrates  other antihypertensives  other drugs for IHD  2 antihypertensives  Betablocker +1 antihypertensive1  3 antihypertensives  Betablocker +2 antihypertensives1  Betablocker +3 antihypertensives1  4 antihypertensives  betablocker+4 antihypertensives  other antihypertensive +antihypertensives1  nitrates +any hypertensive  other drugs for IHD and any antihypertensive and/or nitrate  other antiarrhythmics +any antihypertensives | 321 (27.2)  77 (6.5)  132 (11.2)  55 (4.7)  29 (2.5)  1 (0.1)  0 (0.0)  2 (0.2)  177 (15.0)  92 (8.1)  50 (4.2)  95 (8.1)  25 (2.1)  2 (0.2)  2 (0.2) 9 (0.8)  49 (4.2)  5 (0.4)  57 (4.8) | 9200 (44.2)  1184 (5.7)  2683 (12.9)  773 (3.7)  559 (2.7)  18 (0.1)  12 (0.1)  21 (0.1)  2696 (12.9)  1069 (5.1)  548 (2.6)  975 (4.7)  265 (1.3)  18 (0.1)  19 (0.1)  87 (0.4)  331 (1.6)  15 (0.1)  364 (1.7) |
| Cholesterol-lowering drugs  none  statins  other antilipids  Statins +other antilipids | 708 (60.0)  457 (38.7)  7 (0.6)  8 (0.7) | 14719 (70.6)  5866 (28.2)  135 (0.6)  117 (0.6) |
| Systemic steroids | 123 (10.4) | 1149 (5.5) |
| Antirheumatics  none  disease-modifying antirheumatic drugs  other antirheumatics | 1143 (96.9)  37 (3.1)  0 (0.0) | 20388 (97.8)  446 (2.1)  3 (0.0) |
| Respiratory drugs  none  SABA  LABA/LAMA  inhalation steroid only  SABA +Ipratropium (+/- others)  LABA +steroids  LABA +LAMA +steroids  LAMA +steroids  LABA+LAMA  other pulmonary drugs  other pulmonary drugs +steroids  SABA + LABA/LAMA -steroids  SABA +LABA/LAMA +steroids | 1000 (84.7)  13 (1.1)  19 (1.6)  8 (0.7)  6 (0.5)  45 (3.8)  19 (1.6)  0 (0.0)  7 (0.6)  3 (0.3)  9 (0.8) 6 (0.5)  45 (3.8) | 18754 (90.0)  276 (1.3)  217 (1.0)  211 (1.0)  18 (0.1)  474 (2.3) 122 (0.6)  11 (0.1) 80 (0.4)  32 (0.2)  98 (0.5) 96 (0.5) 448 (2.2) |
| Psychotropic drugs  none  SSRI/SNRI/NaRI  other antidepressants  antipsychotics  benzodiazepines2  anti-cholinergics or memantine  anti-ADHD drugs  NaSSA  other psychotropics  SSRI + other antidepressants  SSRI + NaSSA  SRRI + antipsychotics  SRRI + other psychotropics  benzodiazepines + any psychotropic  antipsychotics + any psychotropic  anti-ADHD + any psychotropic  NaSSA + any psychotropic  other psychotropics + any specified psychotropic | 952 (80.7)  100 (8.5)  1 (0.1)  8 (0.7)  0 (0.0)  6 (0.5)  1 (0.1)  25 (2.1)  28 (2.4)  4 (0.3)  8 (0.7)  11 (0.9) 7 (0.6) 3 (0.3)  20 (1.7)  0 (0.0) 4 (0.3)  2 (0.2) | 18657 (89.5)  1164 (5.6)  17 (0.1) 116 (0.6) 7 (0.0) 27 (0.1)  10 (0.0)  184 (0.9)  182 (0.9) 6 (0.0)  94 (0.5)  87 (0.4)  84 (0.4)  12 (0.1)  149 (0.7)  14 (0.1)  18 (0.1)  9 (0.0) |
| VKA: vitamin K antagonists ASA: Acetylic salicylic acid DOAC: direct oral anticoagulant ADP: Adenosine diphosphate antagonist ACE: angiotensin converting enzyme IHD: Ischemic heart disease SABA: Short-acting beta agonist LABA: long-acting beta agonist LAMA: Long-acting muscarinic antagonist SSRI: Selective serotonin inhibitor SNRI: Serotonin and norepinephrine reuptake inhibitor NaRI: Norepinephrine reuptake inhibitor NaSSA: Norepinephrine and specific serotonergic antidepressants  1either diuretics, ACE/ANG-II inhibitors or Ca2+antagonists 2 likely underreported due to limited general reimbursement for benzodiazepines in Denmark | | |

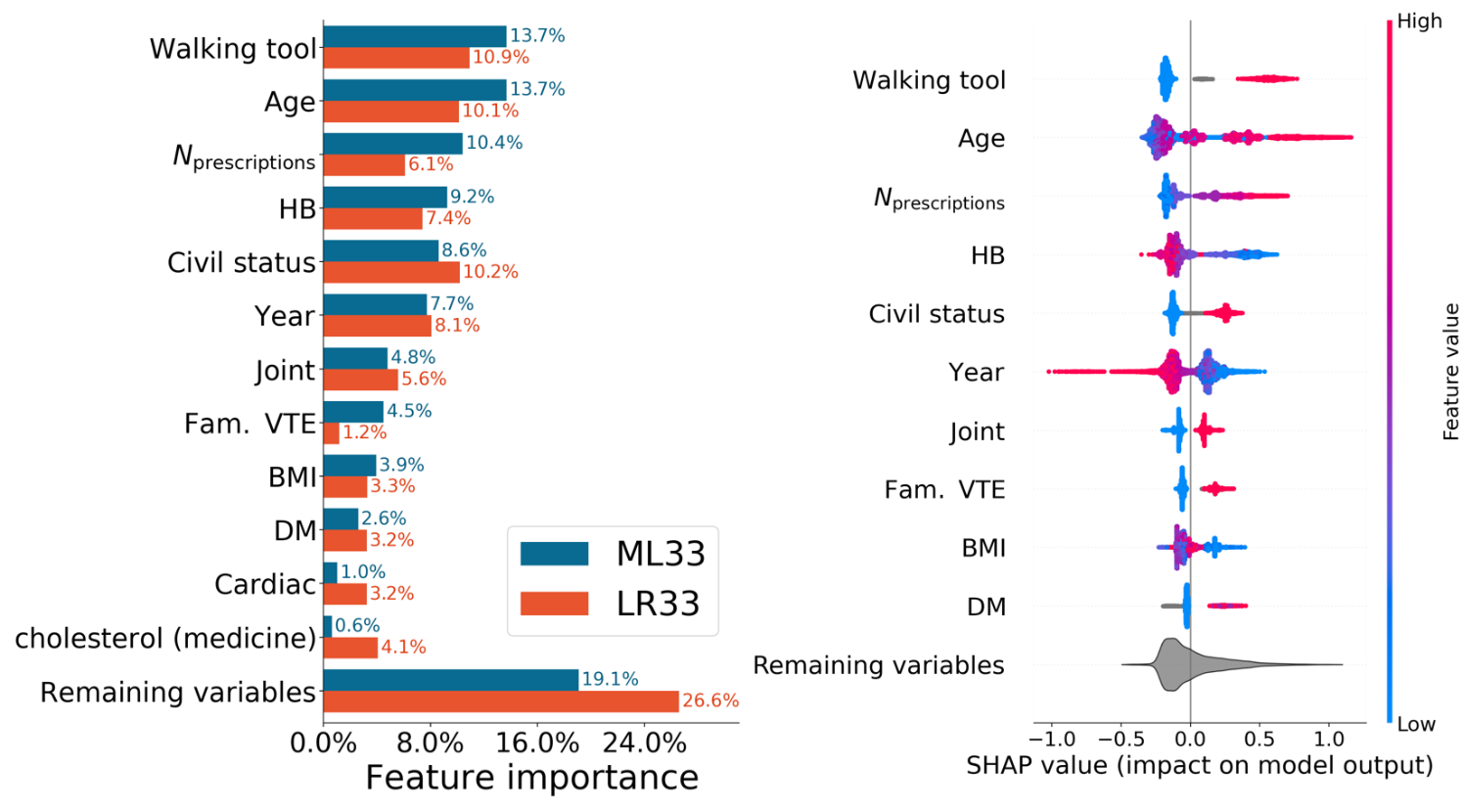
|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table S1: Performance of different models for Outcome B | | | | | | | | | | |
| PPF 20 | TP | FP | FN | TN | TPR | PPV | MCC | AUROC | AUPRC | P (TPR) |
| ML33 | 121 | 661 | 108 | 3023 | 52.8% | 15.5% | 20.5% | 75.3% | 17.1% | - |
| LR33 | 115 | 667 | 114 | 3017 | 50.2% | 14.7% | 18.9% | 74.1% | 16.7% | 28.3% |
| ML10 | 111 | 671 | 118 | 3013 | 48.4% | 14.2% | 17.8% | 74.4% | 16.8% | 17.2% |
| LR10 | 109 | 673 | 120 | 3011 | 47.6% | 13.9% | 17.2% | 73.1% | 16.8% | 12.9% |
| ML-NoAge | 110 | 672 | 119 | 3012 | 48.0% | 14.1% | 17.5% | 72.8% | 16.9% | 15.1% |
| Age\_only | 102 | 661 | 127 | 3023 | 44.5% | 13.4% | 15.8% | 68.7% | 13.4% | 3.8% |
| PPF 25 |  |  |  |  |  |  |  |  |  |  |
| ML33 | 140 | 838 | 89 | 2846 | 61.1% | 14.3% | 20.8% | 75.3% | 17.1% | - |
| LR33 | 136 | 842 | 93 | 2842 | 59.4% | 13.9% | 19.8% | 74.1% | 16.7% | 35.3 |
| ML10 | 134 | 844 | 95 | 2840 | 58.5% | 13.7% | 19.3% | 74.4% | 16.8% | 28.3 |
| LR10 | 125 | 853 | 104 | 2831 | 54.6% | 12.8% | 17.0% | 73.1% | 16.8% | 7.8 |
| ML-NoAge | 121 | 857 | 108 | 2827 | 52.8% | 12.4% | 16.0% | 72.8% | 16.9% | 3.6 |
| Age\_only | 113 | 805 | 116 | 2879 | 49.3% | 12.3% | 15.2% | 68.7% | 13.4% | 0.5 |
| PPF 30 |  |  |  |  |  |  |  |  |  |  |
| ML33 | 153 | 1020 | 76 | 2664 | 66.8% | 13.0% | 20.0% | 75.3% | 17.1% | - |
| LR33 | 147 | 1026 | 82 | 2658 | 64.2% | 12.5% | 18.6% | 74.1% | 16.7% | 27.9 |
| ML10 | 147 | 1026 | 82 | 2658 | 64.2% | 12.5% | 18.6% | 74.4% | 16.8% | 27.7 |
| LR10 | 145 | 1028 | 84 | 2656 | 63.3% | 12.4% | 18.1% | 73.1% | 16.8% | 21.6 |
| ML-NoAge | 140 | 1033 | 89 | 2651 | 61.1% | 11.9% | 17.0% | 72.8% | 16.9% | 10.2 |
| Age\_only | 122 | 933 | 107 | 2751 | 53.3% | 11.6% | 14.8% | 69.8% | 13.4% | 0.1 |
| PPF: positive prediction fraction TP: true positives FP: false positives FN: false negatives TN: true negatives TPR: true positive rate PPV: positive predictive value MCC: Matthews correlation coefficient AURC: area under the ROC curve AUPRC: area under the precision recall curve P(TPR): probability that the model performs better than the ML model relative to TPR. Green/red colors indicates the model with the best/worst performance given that specific metric | | | | | | | | | | |

## Supplemental Digital Content

Figure S1 a+b



S1a) Distribution of risk scores for patients +/- outcome B. The dashed line shows the classification threshold of 20% positive prediction fraction. Figure S1b) ROC curves for the complete machine learning model (ML33), complete logistic regression model (LR33), parsimonious machine learning model (ML10), parsimonious logistic regression model (LR10), machine learning using without age (ML-NoAge) and the age only model (Age).

Figure S2a+b

S2a) The overall importance of the 10 most important variables measured by the SHAP values for both the ML33 and LR33 models. The contributions of the remaining variables are summed in the bottom bar. S2b) The SHAP values for the ML33 model. Positive SHAP-values increase the ML33 risk score while negative values decrease ML33 risk score. The color is related to the value of the variable with blue being lowest and red highest.

Figure S3a-d



SHAP beeswarm plot on the contributions to the ML33 model of individual types of prescribed drugs stratified by age four outcome B.

3a) Anticoagulants 3b) Cardiac drugs 3c) Psychotropics 3d) Respiratory drugs (der er en fejl her vi er på den). NB stofgrupperne kommer på snarest)

**References**

1. Petersen PB, Kehlet H, Jørgensen CC: Improvement in fast-track hip and knee arthroplasty: a prospective multicentre study of 36,935 procedures from 2010 to 2017. Sci Rep 2020; 10: 21233

2. Khan SK, Malviya A, Muller SD, Carluke I, Partington PF, Emmerson KP, Reed MR: Reduced short-term complications and mortality following Enhanced Recovery primary hip and knee arthroplasty: results from 6,000 consecutive procedures. Acta Orthop. 2014; 85: 26-31

3. Savaridas T, Serrano-Pedraza I, Khan SK, Martin K, Malviya A, Reed MR: Reduced medium-term mortality following primary total hip and knee arthroplasty with an enhanced recovery program. Acta Orthop. 2013; 84: 40-3

4. Partridge T, Jameson S, Baker P, Deehan D, Mason J, Reed MR: Ten-Year Trends in Medical Complications Following 540,623 Primary Total Hip Replacements from a National Database. J Bone Joint Surg Am 2018; 100: 360-367

5. Jorgensen CC, Gromov K, Petersen PB, Kehlet H, Lundbeck Foundation Centre for Fast-track H, Knee Replacement Collaborative G: Influence of day of surgery and prediction of LOS > 2 days after fast-track hip and knee replacement. Acta Orthop 2021; 92: 170-175

6. Jorgensen CC, Petersen MA, Kehlet H: Preoperative prediction of potentially preventable morbidity after fast-track hip and knee arthroplasty: a detailed descriptive cohort study. BMJ Open. 2016; 6: e009813

7. Johns WL, Layon D, Golladay G, Kates S, Scott M, Patel NK: Preoperative Risk Factor Screening Protocols in Total Joint Arthroplasty: A Systematic Review. J Arthroplasty 2020

8. Adhia AH, Feinglass JM, Suleiman LI: What Are the Risk Factors for 48 or More-Hour Stay and Nonhome Discharge After Total Knee Arthroplasty? Results From 151 Illinois Hospitals, 2016-2018. J Arthroplasty 2020; 35: 1466-1473.e1

9. Johnson DJ, Castle JP, Hartwell MJ, D'Heurle AM, Manning DW: Risk Factors for Greater Than 24-Hour Length of Stay After Primary Total Knee Arthroplasty. J Arthroplasty 2020; 35: 633-637

10. Shah A, Memon M, Kay J, Wood TJ, Tushinski DM, Khanna V: Preoperative Patient Factors Affecting Length of Stay following Total Knee Arthroplasty: A Systematic Review and Meta-Analysis. J Arthroplasty 2019; 34: 2124-2165.e1

11. Moonesinghe SR, Mythen MG, Das P, Rowan KM, Grocott MP: Risk stratification tools for predicting morbidity and mortality in adult patients undergoing major surgery: qualitative systematic review. Anesthesiology 2013; 119: 959-981

12. Li Q, Zhong H, Girardi FP, Poeran J, Wilson LA, Memtsoudis SG, Liu J: Machine Learning Approaches to Define Candidates for Ambulatory Single Level Laminectomy Surgery. Global Spine J 2021: 2192568220979835

13. Chiew CJ, Liu N, Wong TH, Sim YE, Abdullah HR: Utilizing Machine Learning Methods for Preoperative Prediction of Postsurgical Mortality and Intensive Care Unit Admission. Ann Surg 2020; 272: 1133-1139

14. Li H, Jiao J, Zhang S, Tang H, Qu X, Yue B: Construction and Comparison of Predictive Models for Length of Stay after Total Knee Arthroplasty: Regression Model and Machine Learning Analysis Based on 1,826 Cases in a Single Singapore Center. J Knee Surg 2020

15. Lu Y, Khazi ZM, Agarwalla A, Forsythe B, Taunton MJ: Development of a Machine Learning Algorithm to Predict Nonroutine Discharge Following Unicompartmental Knee Arthroplasty. J Arthroplasty 2021; 36: 1568-1576

16. Shah AA, Devana SK, Lee C, Kianian R, van der Schaar M, SooHoo NF: Development of a Novel, Potentially Universal Machine Learning Algorithm for Prediction of Complications After Total Hip Arthroplasty. J Arthroplasty 2021; 36: 1655-1662.e1

17. Sniderman J, Stark RB, Schwartz CE, Imam H, Finkelstein JA, Nousiainen MT: Patient Factors That Matter in Predicting Hip Arthroplasty Outcomes: A Machine-Learning Approach. J Arthroplasty 2021; 36: 2024-2032

18. Kugelman DN, Teo G, Huang S, Doran MG, Singh V, Long WJ: A Novel Machine Learning Predictive Tool Assessing Outpatient or Inpatient Designation for Medicare Patients Undergoing Total Hip Arthroplasty. Arthroplast Today 2021; 8: 194-199

19. Mohammadi R, Jain S, Namin AT, Scholem Heller M, Palacholla R, Kamarthi S, Wallace B: Predicting Unplanned Readmissions Following a Hip or Knee Arthroplasty: Retrospective Observational Study. JMIR Med Inform 2020; 8: e19761

20. Ramkumar PN, Navarro SM, Haeberle HS, Karnuta JM, Mont MA, Iannotti JP, Patterson BM, Krebs VE: Development and Validation of a Machine Learning Algorithm After Primary Total Hip Arthroplasty: Applications to Length of Stay and Payment Models. J Arthroplasty 2019; 34: 632-637

21. Ramkumar PN, Karnuta JM, Navarro SM, Haeberle HS, Iorio R, Mont MA, Patterson BM, Krebs VE: Preoperative Prediction of Value Metrics and a Patient-Specific Payment Model for Primary Total Hip Arthroplasty: Development and Validation of a Deep Learning Model. J Arthroplasty 2019; 34: 2228-2234.e1

22. Haeberle HS, Helm JM, Navarro SM, Karnuta JM, Schaffer JL, Callaghan JJ, Mont MA, Kamath AF, Krebs VE, Ramkumar PN: Artificial Intelligence and Machine Learning in Lower Extremity Arthroplasty: A Review. J Arthroplasty 2019; 34: 2201-2203

23. Johannesdottir KB, Kehlet H, Petersen BP, Aasvang EK, Sørensen BD, Jorgensen C: Machine learning classifiers do not improve prediction of hospitalization > 2 days after fast-track hip and knee arthroplasty compared

with a classical statistical risk model. acta Orthop; in press

24. Johannesdottir SA, Horvath-Puho E, Ehrenstein V, Schmidt M, Pedersen L, Sorensen HT: Existing data sources for clinical epidemiology: The Danish National Database of Reimbursed Prescriptions. Clin.Epidemiol. 2012; 4: 303-313

25. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS: Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015; 162: W1-73

26. Olczak J, Pavlopoulos J, Prijs J, Ijpma FFA, Doornberg JN, Lundström C, Hedlund J, Gordon M: Presenting artificial intelligence, deep learning, and machine learning studies to clinicians and healthcare stakeholders: an introductory reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. Acta Orthop 2021: 1-13

27. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T: LightGBM: a highly efficient gradient boosting decision tree, Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA, Curran Associates Inc, 2017, pp 3149-57

28. Lin T-Y, Goyal P, Girshick R, He K, Dollár P: Focal Loss for Dense Object Detection. <http://arxiv.org/abs/1708.02002>, ArXiv170802002 Cs 2018

29. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee SI: From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell 2020; 2: 56-67

30. Lundberg SMLSI: A Unified Approach to Interpreting Model Predictions. Edited by Guyon I. Adv Neural Inf Process Syst [Internet], Curran Associates, Inc., 2017

31. McIsaac DI, Wong CA, Bryson GL, van Walraven C: Association of Polypharmacy with Survival, Complications, and Healthcare Resource Use after Elective Noncardiac Surgery: A Population-based Cohort Study. Anesthesiology 2018; 128: 1140-1150

32. Chicco D: Ten quick tips for machine learning in computational biology. BioData Min 2017; 10: 35

33. Chicco D, Tötsch N, Jurman G: The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData Min 2021; 14: 13

34. Tötsch N, Hoffmann D: Classifier uncertainty: evidence, potential impact, and probabilistic treatment. PeerJ Comput Sci 2021; 7: e398

35. Lopez CD, Gazgalis A, Boddapati V, Shah RP, Cooper HJ, Geller JA: Artificial Learning and Machine Learning Decision Guidance Applications in Total Hip and Knee Arthroplasty: A Systematic Review. Arthroplast Today 2021; 11: 103-112

36. Han C, Liu J, Wu Y, Chong Y, Chai X, Weng X: To Predict the Length of Hospital Stay After Total Knee Arthroplasty in an Orthopedic Center in China: The Use of Machine Learning Algorithms. Front Surg 2021; 8: 606038

37. Ramkumar PN, Karnuta JM, Navarro SM, Haeberle HS, Scuderi GR, Mont MA, Krebs VE, Patterson BM: Deep Learning Preoperatively Predicts Value Metrics for Primary Total Knee Arthroplasty: Development and Validation of an Artificial Neural Network Model. J Arthroplasty 2019; 34: 2220-2227.e1

38. Wei C, Quan T, Wang KY, Gu A, Fassihi SC, Kahlenberg CA, Malahias MA, Liu J, Thakkar S, Gonzalez Della Valle A, Sculco PK: Artificial neural network prediction of same-day discharge following primary total knee arthroplasty based on preoperative and intraoperative variables. Bone Joint J 2021; 103-b: 1358-1366

39. Griffiths R, Beech F, Brown A, Dhesi J, Foo I, Goodall J, Harrop-Griffiths W, Jameson J, Love N, Pappenheim K, White S: Peri-operative care of the elderly 2014: Association of Anaesthetists of Great Britain and Ireland. Anaesthesia 2014; 69 Suppl 1: 81-98

40. Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L, Sorensen HT: The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol 2015; 7: 449-90