

BMJ Open

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

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-061753.R2
Article Type:	Original research
Date Submitted by the Author:	n/a
Complete List of Authors:	Michelsen, Christian; University of Copenhagen, The Niel Bohr Institute Jørgensen, Christoffer; Rigshospitalet, Section for Surgical Pathophysiology Heltberg, Mathias; University of Copenhagen, The Niels Bohr Institute Jensen, Mogens; University of Copenhagen, The Niels Bohr institute Lucchetti, Alessandra; University of Copenhagen, The Niels Bohr Institute Petersen, Pelle; Rigshospitalet, Section for Surgical Pathophysiology Petersen, Troels; Kobenhavns Universitet Niels Bohr Instituttet Kehlet, Henrik; Rigshospitalet, Section for Surgical Pathophysiology
Primary Subject Heading:	Anaesthesia
Secondary Subject Heading:	Surgery
Keywords:	Knee < ORTHOPAEDIC & TRAUMA SURGERY, Hip < ORTHOPAEDIC & TRAUMA SURGERY, Orthopaedic & trauma surgery < SURGERY, Adult anaesthesia < ANAESTHETICS, Anaesthesia in orthopaedics < ANAESTHETICS

SCHOLARONE™
Manuscripts

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

Authors:

Christian Michelsen, M.Sci.¹, Christoffer C Jørgensen, M.D., DMSc.^{2,3}, Mathias Heltberg, M.Sci,¹ Mogens H. Jensen, D.Sci., Prof.,¹ Alessandra Lucchetti, M.Sci.,¹ Pelle B Petersen, M.D., Ph.D,^{2,3} Troels Petersen, M.Sci, Ass.Prof.,¹ Henrik Kehlet, M.D., Ph.D., Prof.,^{2,3} The Center for Fast-track Hip and Knee Replacement Collaborative group*

¹The Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17 2100 Copenhagen, Denmark

²Section of Surgical Pathophysiology 7621, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen, Denmark

³The Centre for Fast-track Hip and Knee Replacement, 7621, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen, Denmark

*The members of the Center for Fast-track Hip and Knee Replacement Collaborative group are co-authors of the study and can be found under the subheading Collaborators.

This is a shared first authorship between MC and CJ.

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

Word Count: 3762

ABSTRACT

Objectives: Machine-learning models may improve prediction of length of stay (LOS) and morbidity after surgery. However, few studies include fast-track programs, and most rely on administrative coding with limited follow-up and information on perioperative care. This study investigates benefits of machine-learning models for prediction of postoperative morbidity in fast-track total hip (THA) and knee arthroplasty (TKA).

Design: Cohort study with prospective recording of comorbidity and prescribed medication. Information on length of stay and readmissions through the Danish National Patient Registry and medical records.

Participants: Consecutive unselected primary THA or TKAs between 2014-2017 from seven Danish centers with established fast-track protocols. Data from 2014-2016 (n:18013) was used for training and data from 2017 (n:3913) was used for testing.

Outcomes: Ability of a machine-learning model based on boosted decision trees with 33 preoperative variables for predicting “medical” morbidity leading to LOS >4 days or 90-days readmissions vs. a logistic regression model. We also evaluated a parsimonious machine-learning and logistic regression model using the ten most important variables. Model performances were analyzed using precision, area under receiver operating (AUROC) and precision recall curves (AUPRC) among other performance measures. Variable importance was analyzed using Shapley Additive Explanations values.

Results: Using a threshold of 20% “risk-patients” (n:782), precision, AUROC and AUPRC were 13.6%, 76.3% and 15.5% vs. 12.4%, 74.7% and 15.6% for the machine-learning and logistic regression model, respectively. The parsimonious machine-learning model performed better than the full logistic regression model. Of the top ten variables, eight were shared between the machine-learning and logistic regression models, but with a considerable age-related variation in importance of specific types of medication.

Conclusion: Machine-learning algorithms using preoperative characteristics and prescriptions slightly improved identification of patients in high-risk of “medical” complications after fast-track THA and TKA. Such algorithms could help identify patients who benefit from intensified perioperative care.

STRENGTHS AND LIMITATIONS

Strengths

- Fully implemented fast-track protocols with complete follow-up through nationwide registries and medical records.
- State of the art machine-learning techniques
- Novel analysis on the importance of preoperative prescriptions in predicting postoperative morbidity.

Limitations

- Limited amount of multilevel continuous data, potentially limiting full realization of the machine-learning model.
- Registration of preoperative prescriptions dependent on reimbursement and lack of information on actual use on day of surgery

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 (THA) and knee arthroplasty (TKA), the introduction of enhanced recovery or fast-track programs has led to a significant reduction of postoperative length of stay (length of stay) as well as morbidity and mortality.¹⁻³ However, despite such progress, a fraction of patients still have postoperative complications leading to prolonged length of stay or readmissions.^{1 3 4}

Consequently, in order to prioritize perioperative care, many efforts have been published to preoperatively predict length of stay and morbidity using traditional risk factors such as age, preoperative cardio-pulmonary disease, anemia, diabetes, frailty, etc.⁴⁻⁸ 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 has had a relatively limited clinically relevant ability to predict and reduce potentially preventable morbidity and length of stay.⁴⁻⁸

More recently, machine-learning 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 conditions.^{9 10} This is also the case in THA, TKA and uni-compartmental knee replacement, where several publications on machine-learning algorithms for prediction of length of stay,^{11 12} complications,¹³ disability,¹⁴ potential outpatient setup,¹⁵ readmissions¹⁶ or payment models,^{17 18} have shown promising predictive value compared to conventional statistical methods.¹⁹

However, few papers have included fast-track 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. In our previous study of 9512 THA and TKAs within a fully implemented fast-track protocol and including the above information, we did not find advantages of machine-learning methods compared to logistic regression in predicting a length of stay > 2 days.²⁰ However, this may have been due to data imbalance, lack of details on medication and the chosen outcome of length of stay of >2 days.²⁰ Thus, machine-learning 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 used a large consecutive cohort of patients undergoing fast-track total hip and knee replacement within a national public health-care system¹ to develop an improved

1
2
3 machine-learning model for preoperative prediction of “medical” complications resulting in
4 prolonged length of stay and readmissions. Model performances were subsequently compared
5 to a traditional logistic regression model. In addition to well-defined patient-reported
6 preoperative risk-factors, we also included information on dispensed reimbursed prescriptions 6
months prior to surgery using a nationwide registry.²¹

11
12 **METHODS**
13

14 Reporting of the study is done in accordance with the Transparent reporting of multivariable
15 prediction model for individual prognosis or diagnosis (TRIPOD) statement²² and the Clinical AI
16 Research (CAIR) checklist proposal.²³

17 The study is based on the Centre for Fast-track Hip and Knee Replacement database which is a
18 prospective database on preoperative patient characteristics and enrolling consecutive patients
19 from 7 departments between 2010 and 2017. The database is registered on ClinicalTrials.gov
20 as a study registry (NCT01515670). Patients completed a preoperative questionnaire with nurse
21 assistance if needed. Additional information on reimbursed prescriptions 6 months prior to
22 surgery was acquired using the Danish National Database of Reimbursed Prescriptions
23 (DNDRP) which records all dispensed prescriptions with reimbursement in Denmark.²¹ Finally,
24 data were combined with the Danish National Patient Registry (DNPR) for information on length
25 of stay (counted as postoperative nights spent in hospital), 90-days readmissions with overnight
26 stay and mortality. In case of length of stay >4 days or readmission, patient discharge
27 summaries were reviewed for information on postoperative morbidity and in case of insufficient
28 information, the entire medical records were reviewed. Readmissions were only included if
29 considered related to the surgical procedure, thus excluding planned procedures like cancer
30 workouts, cataract surgery, etc. Readmissions due to urinary tract infection or dizziness after
31 day 30 were also considered unrelated to the surgical procedure. In case of postoperative
32 mortality the entire medical record, including potential readmissions, was reviewed to identify
33 cause of death. Evaluation of discharge and medical records was performed by PP supervised
34 by CJ. In case of disagreement, records were conferred with HK. Subsequently, causes of
35 length of stay >4, readmissions or mortality were classified as “medical” when related to
36 perioperative care (renal failure, falls, pain, thrombosis, anemia, venous thromboembolism or
37 infection etc.) and “surgical” if related to surgical technique (prosthetic infection, revision
38 surgery, periprosthetic fracture, hip dislocation, etc.).¹ In case of a length of stay 4-6 days with a
39 standard discharge summary describing a successful postoperative course, it was assumed that
40 no clinically relevant postoperative complications had occurred. If length of stay was >6 days
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 but with standard discharge summary, the entire medical record was evaluated to confirm that
4 no relevant complications had occurred.
5

6 For the present study, only cases between 2014 and 2017 were used to provide the most up-to
7 date data. All patients had elective unilateral total hip and knee replacement in dedicated
8 arthroplasty departments with similar fast-track protocols, including multimodal opioid sparing
9 analgesia with high-dose (125mg) methylprednisolone, preference for spinal anesthesia, only in-
10 hospital thromboprophylaxis when length of stay \leq 5 days, early mobilization, functional
11 discharge criteria and discharge to own home.¹ There were no selection criteria for the fast-track
12 protocol as it is considered standard of care, but we excluded patients with previous major hip
13 or knee surgery within 90-days of THA or TKA and THA due to severe congenital joint disorder
14 or cancer (Supplemental Material 1).
15
16
17
18
19

20 21 **Patient and Public Involvement** 22

23 There was no involvement of patients or the public in the planning or conduction of the study.
24
25

26 **Outcomes** 27

28 The primary outcome was to develop a machine-learning model to predict the occurrence of
29 “medical” complications resulting in a length of stay >4 days or readmission and compare model
30 performance with a traditional logistic regression model (primary outcome). Secondarily, we
31 investigated how inclusion of cases with a length of stay >4 days but no reported “medical”
32 complication as a positive outcome influenced the model (secondary outcome). For both
33 outcomes, we also investigated whether a parsimonious model including only the top ten
34 variables would perform equally well as the full model. All figures and tables in the main text are
35 based on the primary outcome; the corresponding figures for the secondary outcome are
36 reported in the Supplemental Material.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Statistical Analysis

Data consisted of 33 input variables, of which 7 were continuous. All variables were collected prospectively, either through the patient completed questionnaire, through the DNDRP or a combination of both (table 1). Initially we trimmed the dataset by removing 156 patients (1.7%) who were outliers with regards to weight (<30 kg or >250 kg) and height (<100 cm or >210 cm) or where these data were missing. To reduce the risk of overfitting and allow for unbiased evaluation of model performance, data was subsequently split into a training set consisting of 18013 (82.2%) procedures from 2014-2016 and a test set of 3913 (17.8%) procedures from 2017, as is standard in modelling of data with a temporal component.²⁴ (Supplemental Material 1). These sample sizes are larger than the proposed minima of 3656, when assuming the model will explain 20% of the variability.²⁵ The data analysis was performed in Python and is available online at <https://zenodo.org/record/7330268>.

As reference model, we used logistic regression with missing values being handled by multiple imputations. All variables were then normalized to have zero mean and unit standard deviation by subtracting the original mean and dividing by the original standard deviation. In addition, we used boosted decision trees (LightGBM)²⁶ for the machine-learning models, as such methods work well with categorical data and missing values. We used cross entropy as the objective function for the machine-learning model.

The full machine-learning model was trained and hyperparameter optimized using the state of the art framework [Optuna](#)²⁷ with the [Tree-structured Parzen](#) Estimator algorithm²⁸ to efficiently sample hyperparameters and with a median stopping rule to minimize optimization time. The models were trained on the training data and then used for making predictions on the unseen test data (Supplemental Material 1). The classification threshold was calibrated such that 20% of the total number of patients were predicted as positive by the model (positive predictive fraction of 20%). We also included results for values of 25% and 30%. Furthermore, we trained two parsimonious models using machine-learning and logistic regression with only the 10 most important features. All mentioned models were calibrated using Platt's method (Supplemental Material 2).²⁹ Finally, we constructed a model based on age alone (Age) to explore the added value of multiple variable prediction.

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.^{30 31} Finally, we investigated a potential relation between reimbursed prescribed cardiac drugs, anticoagulants, psychotropics and pulmonary drugs and age, as the

relation between polypharmacy and postoperative outcomes have mainly been found in older patients.³²

For evaluating model performance, we computed the number of true positives (TP), false positives (FP), false negatives (FN), true negatives (TN), sensitivity (true positive rate = TP / (TP+FN)), precision (positive predictive value = TP / (TP+FP)). Since the data was quite imbalanced (about a 1:20 positive:negative ratio) we also computed the Matthews Correlation Coefficient (MCC) which is independent of class imbalance.^{33 34} 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 $P(\text{sensitivity})$,³⁵ which is the probability that a model will perform better than the machine-learning model relative to the sensitivity. Thus, for two equally performing models $P(\text{sensitivity})$ is $\approx 50\%$.

RESULTS

Median age in the 3913 patients was 70 years (IQR 62-76), 59% were female and 58% had THA (table 1).

Table 1. patient demographics with and without the primary outcome (length of stay >4 days or readmissions due to “medical” morbidity) in the combined test and training dataset.

Preoperative characteristics	training set (n:18013)	test set (n:3913)
n (%) unless otherwise specified		
mean age (SD)	69.0 (62.0-75.0)	70.0 (62.0-76.0)
mean number of reimbursed prescriptions ¹ (SD)	2.0 (0.0-3.0)	2.0 (0.0-3.0)
female gender	755 (64.0)	12133 (58.2)
hip arthroplasty	9918 (54.8)	2260 (57.8)
mean weight in kg (SD)	80.5 (70.0-93.0)	81.0 (70.0-92.0)
mean height in cm (SD)	170.0 (164.0-177.0)	170.0 (164.0-177.0)
mean body mass index (SD)	27.5 (24.6-31.2)	27.5 (24.6-31.1)
regular use of walking aid		
missing	552 (46.8)	4398 (21.5)
29 (2.5)	359 (1.7)	
living alone	5914 (32.9)	1381 (35.7)
with others	11971 (66.5)	2469 (63.8)
institution		
missing	116 (0.6)	21 (0.5)
12 (0.6)	42 (1.1)	
Hemoglobin (SD)	8.6 (8.1-9.1)	8.6 (8.1-9.2)
Missing	291 (1.5)	55 (1.4)
>2 units of alcohol/day	1382 (7.7)	286 (7.4)
Missing	57 (0.8)	36 (0.9)

active smoker	130 (11.0)	2751 (13.2)
missing	11 (0.9)	141 (0.7)
cardiac disease	2527 (14.0)	529 (13.7)
missing	17 (0.6)	53 (1.4)
hypercholesterolemia	5396 (29.9%)	1133 (29.3%)
missing	83 (0.5)	44 (1.2)
hypertension	9030 (51.4)	1849 (49.5)
missing	546 (3.0)	179 (4.6)
pulmonary disease	1668 (9.2)	355 (9.2)
missing	63 (0.4)	38 (1.0)
previous cerebral attack	1038 (5.8)	213 (5.6)
missing	157 (1.3)	77 (2.0)
previous VTE	1331 (7.5)	283 (7.4)
missing	283 (1.6)	66 (1.7)
malignancy (undefined)	1469 (8.1)	134 (3.4)
previous radically treated malignancy	1752 (9.7)	440 (11.2)
missing	136 (0.8)	40 (1.0)
chronic kidney disease	266 (1.5)	57 (1.5)
missing	276 (1.5)	50 (1.3)
family member with VTE	2235 (14.1)	430 (12.5)
missing	2189 (12.6)	479 (12.2)
regular snoring	266 (22.5)	5522 (26.5)
uncertain about snoring	208 (17.6)	3781 (18.1)
missing	259 (21.9)	3309 (15.9)
not feeling rested	7272 (42.4)	9340 (44.8)
uncertain about being rested	48 (4.1)	809 (3.9)
missing	105 (8.9)	1230 (5.9)
psychiatric disorder	1464 (8.4)	282 (7.6)
missing	580 (3.2)	182 (4.7)

Characteristic based on combination of questionnaire and DNDRP

Diabetes

diet treated diabetes ²	251 (1.4)	52 (1.3)
oral antidiabetics	1294 (7.2)	291 (7.5)
insulin treated diabetes ³	405 (2.2)	68 (1.8)
missing	68 (0.4)	36 (0.9)

SD: standard deviation 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 ³+/- oral antidiabetics

Details on prescribed drug types are shown in Supplemental Material 3. Median length of stay was 2 (IQR: 1-2) days with 7.6% 90-days readmissions and the primary outcome occurring in 182 (4.7%) patients. When applying any model with a positive prediction fraction of 20% to the 3913 patients, 782 qualified as "risk-patients". The results are summarized in figure 1 and table 2.

Table 2: Performance of the models with a predefined positive prediction fraction of 20% for primary outcome.

Positive prediction fraction 20%	TF	FP	FN	TN	Sensitivity %	Precision %	MCC %	AUROC %	AUPRC %	Brier %	P (sensitivity) %
Full machine-learning model	106	676	76	3055	58.2	13.6	21.1	76.3	15.5	4.19	-
Full logistic regression model	97	685	85	3046	53.3	12.4	18.4	74.7	15.6	4.32	17.2
Parsimonious machine-learning model	100	682	82	3049	54.9	12.8	19.3	75.9	17.3	4.34	26.4
Parsimonious logistic regression model	90	692	92	3039	49.5	11.5	16.3	73.8	15.8	4.33	4.86
Age-only model	87	676	95	3055	47.8	11.4	15.8	69.7	12.1	38.8	3.55

TP: true positives FP: false positives FN: false negatives TN: true negatives MCC: Matthews correlation coefficient AUROC: area under the operating receiver curve AUPRC: area under the precision recall curve P(sensitivity): probability that a model performs better than the machine-learning model relative to sensitivity.

When considering risk scores from the full machine-learning (figure 1a) and full logistic regression model leading to this risk-patient selection, 106 and 98 had the primary outcome, respectively. Correspondingly, the sensitivity and precision were 58.2% and 13.6% for the full machine-learning and 53.3% and 12.4% for the full logistic regression model, respectively. The full machine-learning model was superior (figure 1b) on all parameters (except AUPRC) compared to any of the other models, although the differences were minor (table 2).

The results were similar when using positive prediction fractions of 25% and 30%, but with the sensitivity for the full machine-learning model increasing to 64.3% and 69.2% and precision decreasing to 12.0% and 10.7%, respectively (Supplemental Material 4). Despite age being the single most important variable, age alone had a significantly lower sensitivity at 47.8%. When evaluating feature importance, we found a strong correlation between the full machine-learning and full logistic regression model, with age and use of walking aids being the most important variables in both (figure 2a). From the combined importance of variables outside the top ten, the machine-learning approach extracted more information with fewer variables than logistic regression (figure 1b).

For the full machine-learning model, 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, a recent date of surgery and an 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 vitamin-K antagonists (VKA) or adenosine diphosphate (ADP) antagonists increased, while acetylic salicylic acid and direct oral anticoagulants (DOAC) reduced the risk score of the full machine-learning model, regardless of age (figure 3a). The SHAP analysis of prescribed cardiac drugs revealed that prescriptions on Ca^{2+} -antagonists and betablockers in combination with one or two other antihypertensives increased the risk-score, as did prescriptions on nitrates, other antihypertensives and antiarrhythmics. For the remaining cardiac drugs, prescriptions either reduced or had minor influence, and with limited relation with age (figure 3b). Preoperative psychotropic prescriptions increased the risk-score except for antipsychotics (0.6%). For users of selective serotonin inhibitors there was a clear age-related distinction with the risk score being increased in elderly patients but decreased in those < 60 years (figure 3c). Finally, the risk score increased with prescriptions on inhalation steroid and β -blockers, and more accentuated in the younger patients (figure 3d).

The results including patients with a length of stay >4 days, but no reported postoperative complications (secondary outcome) were similar as for the primary outcome. In general, we found that the full machine-learning model was superior to the others, although the differences were smaller than for the primary outcome. (Supplemental Material 5 listing outcome parameters and Supplemental material 6 figure S1a-b showing distributions and ROC curves for the secondary outcome). While the ten most important variables for the full machine-learning model remained unchanged, familiar disposition for venous thromboembolism replaced gender as one of the top ten important variables in the full logistic regression model (Supplemental material 7 figure S2a-b showing SHAP values for the secondary outcome). Furthermore, the SHAP analysis on specific prescribed drugs demonstrated that the machine-learning model found no benefits from information on prescriptions on respiratory drugs, why all SHAP values were zero. In addition, the reduced risk with acetylsalicylic acid and DOAC prescriptions, as well as the influence of practically all cardiac drugs except for nitrates, other antihypertensives and antiarrhythmics, was attenuated (Supplemental material 8 figure S3a-d showing SHAP-values of prescriptions of specific drugs for the secondary outcome).

DISCUSSION

We found that using a machine-learning algorithm including all 33 available variables and a parsimonious machine-learning-algorithm encompassing only the 10 most important predictors

improved prediction of patients at increased risk of having a length of stay >4 days or readmissions due to medical complications compared to traditional logistic regression models. In contrast, when also including patients having a length of stay >4 days but without a well-defined complication as an outcome, the parsimonious machine-learning model was slightly worse than a traditional logistic regression model including all variables. We also found that although age was the single most important predictor of both the primary and the secondary outcome, it was less suited for prediction of postoperative medical complications after fast-track THA and TKA on its own. Finally, we demonstrated how the chosen classification threshold of the machine-learning algorithm influenced model performance through an increase in sensitivity at the cost of decreased precision.

A previous systematic review also found that machine-learning algorithms may provide better prediction of postoperative outcomes in THA and TKA.³⁶ However, the authors concluded that such models performed best at predicting postoperative complications, pain and patient reported outcomes and were less accurate at predicting readmissions and reoperations.³⁶ That machine-learning 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 machine-learning framework to predict selected major complications after THA.¹³ However, theirs 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 found 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 machine-learning studies,³⁷ our dataset included not only preoperative data but also only one paraclinical variable, which 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 machine-learning algorithms for home discharge in uni-compartmental knee replacement¹² and spine surgery,⁹ they are not standard in fast-track protocols and not easy to interpret from a pathophysiological point of view. Most decisions on which patients may benefit from more extensive postoperative care will likely need to be conducted preoperatively, as there is an increasing need to prioritize limited health-care resources. Thus, although postoperative information such as duration of surgery, perioperative blood length of stays or postoperative hemoglobin have been included in other

1
2
3 studies³⁷, we decided against the use of peri- and postoperative data. The same approach has
4 been used by Ramkumar *et al.* who used U.S. National Inpatient Sample data including 15
5 preoperative variables, to predict length of stay, patient charges and disposition after both TKA³⁸
6 and THA.¹⁸ However, these studies were not conducted in a socialized health care system, and
7 the main focus was on the need for differentiated payment bundles and without specific
8 information on the reason for increased length of stay or non-home discharge.³⁸ Wei *et al.* used
9 an artificial neural network model to predict same-day discharge after TKA, based on the
10 NSQUIP database from 2018 and found that six of the ten most important variables were the
11 same compared with logistic regression, similar to our findings.³⁹ However, patients with one-
12 day length of stay were intentionally excluded due to variations in in-patient vs. out-patient
13 registration.³⁹

14
15 Age has traditionally been a major factor when predicting surgical outcomes and remained the
16 single most important predictor in our study.. However, although elderly patients had increased
17 risk of postoperative complications, likely related to decline of physical reserves,⁴⁰ the use of
18 chronological age alone was inferior compared to both machine-learning and logistic regression
19 models incorporating comorbidity and functional status. Thus, using age by itself for identifying
20 the high-risk population resulted in missing 18% of the “true risk-patients” (87 compared to 106
21 in the full ML model).

22 We used the SHAP values for estimation of feature importance, thus providing a better
23 understanding of the otherwise “black-box” machine-learning model. The SHAP values showed
24 which variables contribute most to the risk-score predictions. In this context, inclusion of specific
25 data on reimbursed prescriptions 6 months prior to surgery based upon the unique Danish
26 registries, unsurprisingly found increased risk-scores with increased number of prescriptions
27 and with the majority being in elderly patients. Similarly, a Canadian study in elective non-
28 cardiac surgery found decreased survival and increased length of stay and readmissions and
29 costs in patients >65 years with polypharmacy.³² However, this is a complex relationship where
30 some patients benefit from their treatments, while other may suffer from undesirable side-
31 effects. Consequently, the authors cautioned against altering perioperative practices based on
32 current evidence.³² However, the information from the included prescriptions with SHAP
33 analysis may provide inspiration for new hypothesis-generating studies such as investigation of
34 the potential differences in risk-profile between having preoperative prescribed VKA and
35 DOAKs. Also, the age-related differences in risk from SSRI’s in our study could guide further
36 studies on “deprescription”.
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 Another requirement for machine-learning-algorithms to be clinically useful is user friendliness
4 and not depending on excessive additional data collection by the attending clinicians. In this
5 context, it was a bit disappointing that the parsimonious machine-learning algorithm with only
6 the ten most important variables was slightly worse at predicting the secondary outcome than
7 the full logistic regression model. A reason for this could be that when including a length of stay
8 >4 days but without described medical complications, the combination of all variables provides
9 information not available by merely including the ten most important ones. This highlights the
10 need for as much detailed, and preferably non-binary, data as possible to fulfill the true potential
11 of machine-learning algorithms.
12
13

14 Our study has some limitations. First, one of the strengths of machine learning compared to
15 logistic regression is the analysis of multilevel continuous data, whereas we included only a
16 limited number of, often binary, preoperative variables. This could have limited the full
17 realization of our machine-learning model. As previously mentioned, we excluded intraoperative
18 information, including type of anesthesia, surgical approach etc. all of which may influence
19 postoperative outcomes. The observational design of this study means that we cannot exclude
20 unmeasured confounding or confounding by indication. Also, despite that the DNDRP has a
21 near complete registration of dispensed medicine in Denmark, some types or drugs, especially
22 benzodiazepines, are exempt from general reimbursement and thus not sufficiently captured.²¹
23 Furthermore, it is doubtful whether the patients used all types of drugs at the time of surgery
24 (e.g. heparin which is rarely for long-term use). Finally, classification of a complication being
25 "medical" depended on review of the discharge records which can also introduce bias. However,
26 we believe our approach to be superior to depending only on diagnostic codes which often are
27 inaccurate⁴¹ and provide limited details on whether the complication may be attributed to a
28 medical or surgical adverse event. The strengths of our study include the use of national
29 registries with high degree of completion (>99% of all somatic admissions in case of the
30 DNDRP),⁴² prospective recording of comorbidity, extensive information on prescription patterns
31 6 months prior to surgery and similar established enhanced recovery protocols in all
32 departments.
33
34

35 In summary, our results suggest that machine-learning-algorithms likely provide clinically
36 relevant improved predictions for defining patients in high-risk of medical complications after
37 fast-track THA and TKA compared to a logistic regression model. Future studies could benefit
38 from using such algorithms to find a manageable population of patients who benefit from
39 intensified perioperative care.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

1
2
3 **Competing interests:** Prof. Kehlet is a board member of “Rapid Recovery”, by Zimmer Biomet.
4 Mr. Heltberg is sponsored by a grant from the Lundbeck Foundation, independently of the
5 present study. Dr. Petersen is an advisory member of Sanofi outside of the present study. The
6 remaining authors declare no conflicts of interest.
7
8 **Funding:** The study received funding from the Lundbeck Foundation, Copenhagen, Denmark
9 grant number R25-A2702.

10
11 **Author Contributions:** CJ and HK contributed to the original idea of the study. CJ, PP and HK
12 contributed to data collection and review of medical records. CM, MH, MJ, AL and TP
13 contributed to the statistical methods, designed the prediction models and conducted the
14 statistical analysis. CM, CJ, HK and TP wrote the original draft. All authors contributed to
15 revision of the initial draft and agreed on the final version of the manuscript. The members of the
16 Centre for Fast-track Hip and Knee Replacement Database collaborative group all contributed
17 by implementing the fast-track protocol at their respective departments and reviewing the final
18 manuscript.
19
20

21 **Collaborators:**
22

23 Frank Madsen M.D. Consultant, Department of Orthopedics, Aarhus University Hospital,
24 Aarhus, Denmark
25 Torben B. Hansen M.D., Ph.D., Prof. Department of Orthopedics, Holstebro Hospital, Holstebro,
26 Denmark
27 Kirill Gromov, M.D., Ph.D., Ass.Prof. Department of Orthopedics Hvidovre Hospital, Hvidovre
28 Denmark
29 Thomas Jakobsen, M.D., Ph.D., DM.Sci., Ass. Prof. Department of Orthopedics, Aalborg
30 University Hospital, Farsø, Denmark
31 Claus Varnum, M.D., Ph.D., Ass. Prof. Department of Orthopedic Surgery, Lillebaelt Hospital -
32 Vejle, University Hospital of Southern Denmark, Denmark
33 Soren Overgaard, M.D., DM.Sci., Prof, Department of Orthopedics, Bispebjerg Hospital,
34 Copenhagen, Denmark
35 Mikkel Rathbun, M.D., Ph.D., Ass. Prof. Department of Orthopedics, Gentofte Hospital,
36 Gentofte, Denmark
37 Lars Hansen, M.D., Consultant, Department of Orthopedics, Sydvestjysk Hospital, Grindsted,
38 Denmark
39
40
41
42
43

44 **Data sharing:** The original dataset is not publicly available due to Danish data-protection law
45 but can be acquired from the corresponding author by request. All statistical code can be freely
46 accessed from <https://zenodo.org/record/7330268>
47

48 **Ethics and Permissions**
49

50 No Ethics Committee approval was necessary as the National Danish Ethics committee exempt
51 non-interventional observational studies. Permission to review and store information from
52 medical records without informed consent was acquired from Center for Regional Development
53 (R-20073405) and the Danish Data Protection Agency (RH-2007-30-0623).
54
55
56
57
58
59

References

- Petersen PB, Kehlet H, Jorgensen CC, et al. Improvement in fast-track hip and knee arthroplasty: a prospective multicentre study of 36,935 procedures from 2010 to 2017. *Scientific reports* 2020;10(1):21233.
- Khan SK, Malviya A, Muller SD, et al. Reduced short-term complications and mortality following Enhanced Recovery primary hip and knee arthroplasty: results from 6,000 consecutive procedures. *Acta Orthop* 2014;85(1):26-31.
- 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. *The Journal of bone and joint surgery American volume* 2018;100(5):360-67.
- Jorgensen CC, Gromov K, Petersen PB, et al. Influence of day of surgery and prediction of LOS > 2 days after fast-track hip and knee replacement. *Acta orthopaedica* 2021;92(2):170-75.
- 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(1):e009813.
- Johns WL, Layon D, Golladay GJ, et al. Preoperative Risk Factor Screening Protocols in Total Joint Arthroplasty: A Systematic Review. *J Arthroplasty* 2020;35(11):3353-63.
- 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(6):1466-73 e1.
- Shah A, Memon M, Kay J, et al. Preoperative Patient Factors Affecting Length of Stay following Total Knee Arthroplasty: A Systematic Review and Meta-Analysis. *J Arthroplasty* 2019;34(9):2124-65 e1.
- Li Q, Zhong H, Girardi FP, et al. Machine Learning Approaches to Define Candidates for Ambulatory Single Level Laminectomy Surgery. *Global spine journal* 2021;2192568220979835.
- Chiew CJ, Liu N, Wong TH, et al. Utilizing Machine Learning Methods for Preoperative Prediction of Postsurgical Mortality and Intensive Care Unit Admission. *Annals of surgery* 2020;272(6):1133-39.
- Li H, Jiao J, Zhang S, et al. 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. *The journal of knee surgery* 2022;35(1):7-14.
- Lu Y, Khazi ZM, Agarwalla A, et al. Development of a Machine Learning Algorithm to Predict Nonroutine Discharge Following Unicompartmental Knee Arthroplasty. *J Arthroplasty* 2021;36(5):1568-76.
- Shah AA, Devana SK, Lee C, et al. Development of a Novel, Potentially Universal Machine Learning Algorithm for Prediction of Complications After Total Hip Arthroplasty. *J Arthroplasty* 2021;36(5):1655-62 e1.
- Sniderman J, Stark RB, Schwartz CE, et al. Patient Factors That Matter in Predicting Hip Arthroplasty Outcomes: A Machine-Learning Approach. *J Arthroplasty* 2021;36(6):2024-32.
- Kugelman DN, Teo G, Huang S, et al. A Novel Machine Learning Predictive Tool Assessing Outpatient or Inpatient Designation for Medicare Patients Undergoing Total Hip Arthroplasty. *Arthroplasty today* 2021;8:194-99.
- Mohammadi R, Jain S, Namin AT, et al. Predicting Unplanned Readmissions Following a Hip or Knee Arthroplasty: Retrospective Observational Study. *JMIR medical informatics* 2020;8(11):e19761.

- 1
2
3 17. Ramkumar PN, Navarro SM, Haeberle HS, et al. Development and Validation of a Machine
4 Learning Algorithm After Primary Total Hip Arthroplasty: Applications to Length of Stay
5 and Payment Models. *J Arthroplasty* 2019;34(4):632-37.
- 6 18. Ramkumar PN, Karnuta JM, Navarro SM, et al. Preoperative Prediction of Value Metrics
7 and a Patient-Specific Payment Model for Primary Total Hip Arthroplasty: Development
8 and Validation of a Deep Learning Model. *J Arthroplasty* 2019;34(10):2228-34 e1.
- 9 19. Haeberle HS, Helm JM, Navarro SM, et al. Artificial Intelligence and Machine Learning in
10 Lower Extremity Arthroplasty: A Review. *J Arthroplasty* 2019;34(10):2201-03.
- 11 20. Johannesson KB, Kehlet H, Petersen PB, et al. Machine learning classifiers do not
12 improve prediction of hospitalization > 2 days after fast-track hip and knee arthroplasty
13 compared with a classical statistical risk model. *Acta orthopaedica* 2022;93:117-23.
- 14 21. Johannesson SA, Horvath-Puhó E, Ehrenstein V, et al. Existing data sources for clinical
15 epidemiology: The Danish National Database of Reimbursed Prescriptions.
16 *ClinEpidemiol* 2012;4:303-13.
- 17 22. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable
18 prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and
19 elaboration. *Annals of internal medicine* 2015;162(1):W1-73.
- 20 23. Olczak J, Pavlopoulos J, Prijs J, et al. Presenting artificial intelligence, deep learning, and
21 machine learning studies to clinicians and healthcare stakeholders: an introductory
22 reference with a guideline and a Clinical AI Research (CAIR) checklist proposal. *Acta
23 orthopaedica* 2021;92(5):513-25.
- 24 24. Tashman L. Out-of-Sample tests of forecasting accuracy: an analysis and review.
25 *International Journal of Forecasting* 2000;16(4):437-50.
- 26 25. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a
27 clinical prediction model. *Bmj* 2020;368:m441.
- 28 26. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T. LightGBM: a highly efficient
29 gradient boosting decision tree. Proceedings of the 31st International Conference on
30 Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc,
31 2017:3149-57.
- 32 27. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation
33 Hyperparameter Optimization Framework. Secondary Optuna: A Next-generation
34 Hyperparameter Optimization Framework 2019. <http://arxiv.org/abs/1907.10902>.
- 35 28. Bergstra J, Bardenet R, Bengio Y, et al. Algorithms for Hyper-Parameter Optimization.
36 2011; 24.
37 <https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cf12577bc2619bc635690-Paper.pdf>.
- 38 29. Platt J. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized
39 Likelihood Methods. CiteSeer, 2000.
- 40 30. Lundberg SM, Erion G, Chen H, et al. From Local Explanations to Global Understanding
41 with Explainable AI for Trees. *Nature machine intelligence* 2020;2(1):56-67.
- 42 31. Lundberg SMLS. A Unified Approach to Interpreting Model Predictions. In: Guyon I, ed. Adv
43 Neural Inf Process Syst [Internet]: Curran Associates, Inc., 2017.
- 44 32. McIsaac DI, Wong CA, Bryson GL, et al. Association of Polypharmacy with Survival,
45 Complications, and Healthcare Resource Use after Elective Noncardiac Surgery: A
46 Population-based Cohort Study. *Anesthesiology* 2018;128(6):1140-50.
- 47 33. Chicco D. Ten quick tips for machine learning in computational biology. *BioData mining*
48 2017;10(1):35 (2017).
- 49 34. Chicco D, Totsch N, Jurman G. The Matthews correlation coefficient (MCC) is more reliable
50 than balanced accuracy, bookmaker informedness, and markedness in two-class
51 confusion matrix evaluation. *BioData mining* 2021;14(1):13 (2021).
- 52
53
54
55
56
57
58
59
60

- 1
2
3 35. Totsch N, Hoffmann D. Classifier uncertainty: evidence, potential impact, and probabilistic
4 treatment. *PeerJ Computer science* 2021;7:e398.
- 5 36. Lopez CD, Gazgalis A, Boddapati V, et al. Artificial Learning and Machine Learning Decision
6 Guidance Applications in Total Hip and Knee Arthroplasty: A Systematic Review.
7 *Arthroplasty today* 2021;11:103-12.
- 8 37. Han C, Liu J, Wu Y, et al. To Predict the Length of Hospital Stay After Total Knee
9 Arthroplasty in an Orthopedic Center in China: The Use of Machine Learning Algorithms.
10 *Frontiers in surgery* 2021;8:606038.
- 11 38. Ramkumar PN, Karnuta JM, Navarro SM, et al. Deep Learning Preoperatively Predicts
12 Value Metrics for Primary Total Knee Arthroplasty: Development and Validation of an
13 Artificial Neural Network Model. *J Arthroplasty* 2019;34(10):2220-27 e1.
- 14 39. Wei C, Quan T, Wang KY, et al. Artificial neural network prediction of same-day discharge
15 following primary total knee arthroplasty based on preoperative and intraoperative
16 variables. *Bone Joint J* 2021;103-B(8):1358-66.
- 17 40. Griffiths R, Beech F, Brown A, et al. Peri-operative care of the elderly. *Anaesthesia* 2014;69
18 Suppl 1:81-98.
- 19 41. Bedard NA, Pugely AJ, McHugh MA, et al. Big Data and Total Hip Arthroplasty: How Do
20 Large Databases Compare? *J Arthroplasty* 2018;33(1):41-45.e3.
- 21 42. Schmidt M, Schmidt SA, Sandegaard JL, et al. The Danish National Patient Registry: a
22 review of content, data quality, and research potential. *Clinical epidemiology* 2015;7:449-
23 90.
- 24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

FIGURE LEGENDS**Figure 1a-b**

1a) Distribution of full machine learning model risk scores for patients +/- the primary outcome.
The dashed line marks the classification threshold of 20% positive prediction fraction.

1b) Receiver operating curves (ROC) for the full machine learning model (F-MLM), full logistic regression model (F-LRM), parsimonious machine learning model (P-MLM), parsimonious logistic regression model (P-LRM) and the age-only model (AM).

Figure 2a-b

2a) The overall importance of the 10 most important variables measured by the SHAP-values for the full machine-learning and full logistic regression models on the primary outcome (LOS >4 days or readmission due to "medical" morbidity). Only the importance of prescribed anticholesterols and gender differ between the models. The contributions of the remaining variables are summed in the bottom bar.

2b) The SHAP-values for the full machine-learning model on the primary outcome, where positive increase and negative values decrease the risk score. Each dot represents a patient and the color is related to the value of the variable with blue being lowest and red highest..

Figure 3a-d

SHAP scatter-plot on the contributions to the full machine-learning model on the primary outcome (LOS >4 days or readmission due to "medical" morbidity), for individual types of prescribed anticoagulants, cardiac drugs, psychotropics and respiratory drugs stratified by age.

3a) Prescribed anticoagulants

VKA: vitamin K antagonists ASA: acetylsalicylic acid DOAC: direct oral anticoagulant ADP: Adenosine diphosphate ACE: angiotensin converting enzyme

3b) Prescribed cardiac drugs

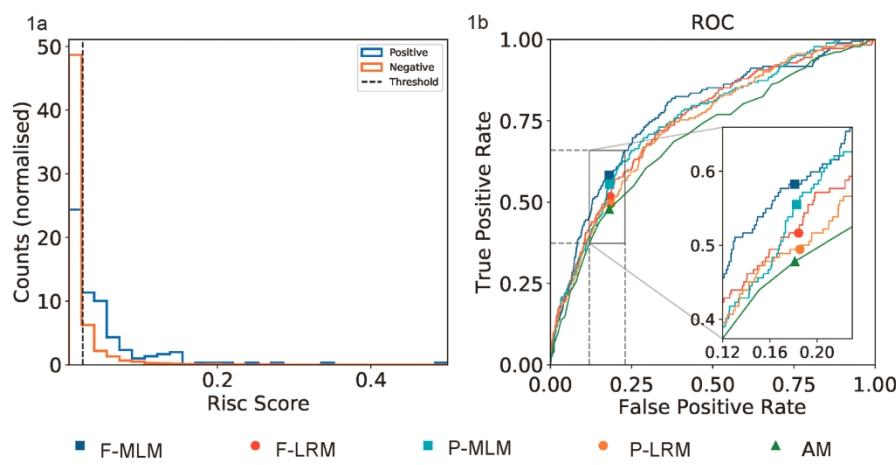
ACE: angiotensin converting enzyme AHT: antihypertensive. Other AHT were defined as AHT different from diuretics ANG-II/ACE inhibitors or Ca²⁺antagonists. IHD: Ischemic heart disease

3c) Prescribed psychotropics

SSRI: Selective serotonin inhibitor SNRI: Serotonin and norepinephrine reuptake inhibitor NaRI: Norepinephrine reuptake inhibitor NaSSA: Norepinephrine and specific serotonergic antidepressants. AD: antidepressants BZ: Benzodiazepines (likely underreported due to limited general reimbursement in Denmark). ADHD: Attention-deficit/hyperactivity disorder

3d) Prescribed respiratory drugs

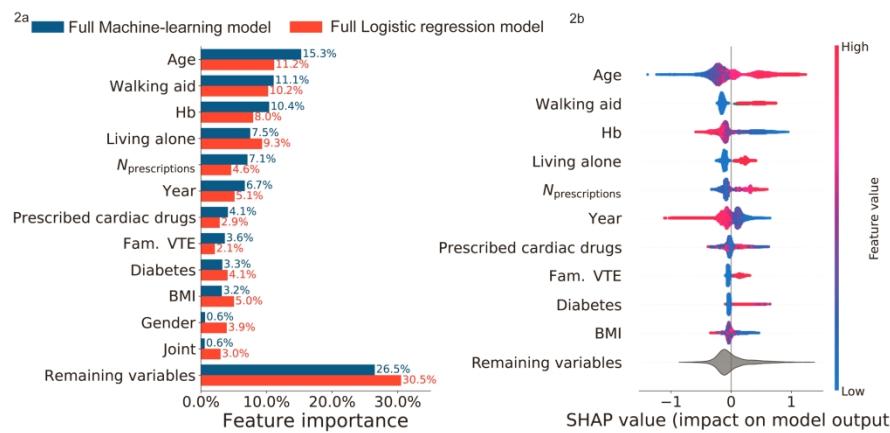
SABA: Short-acting beta agonist LABA: long-acting beta agonist LAMA: Long-acting muscarinic antagonist.



1a) Distribution of full machine learning model risk scores for patients +/- the primary outcome. The dashed line marks the classification threshold of 20% positive prediction fraction.

1b) Receiver operating curves (ROC) for the full machine learning model (F-MLM), full logistic regression model (F-LRM), parsimonious machine learning model (P-MLM), parsimonious logistic regression model (P-LRM) and the age-only model (AM).

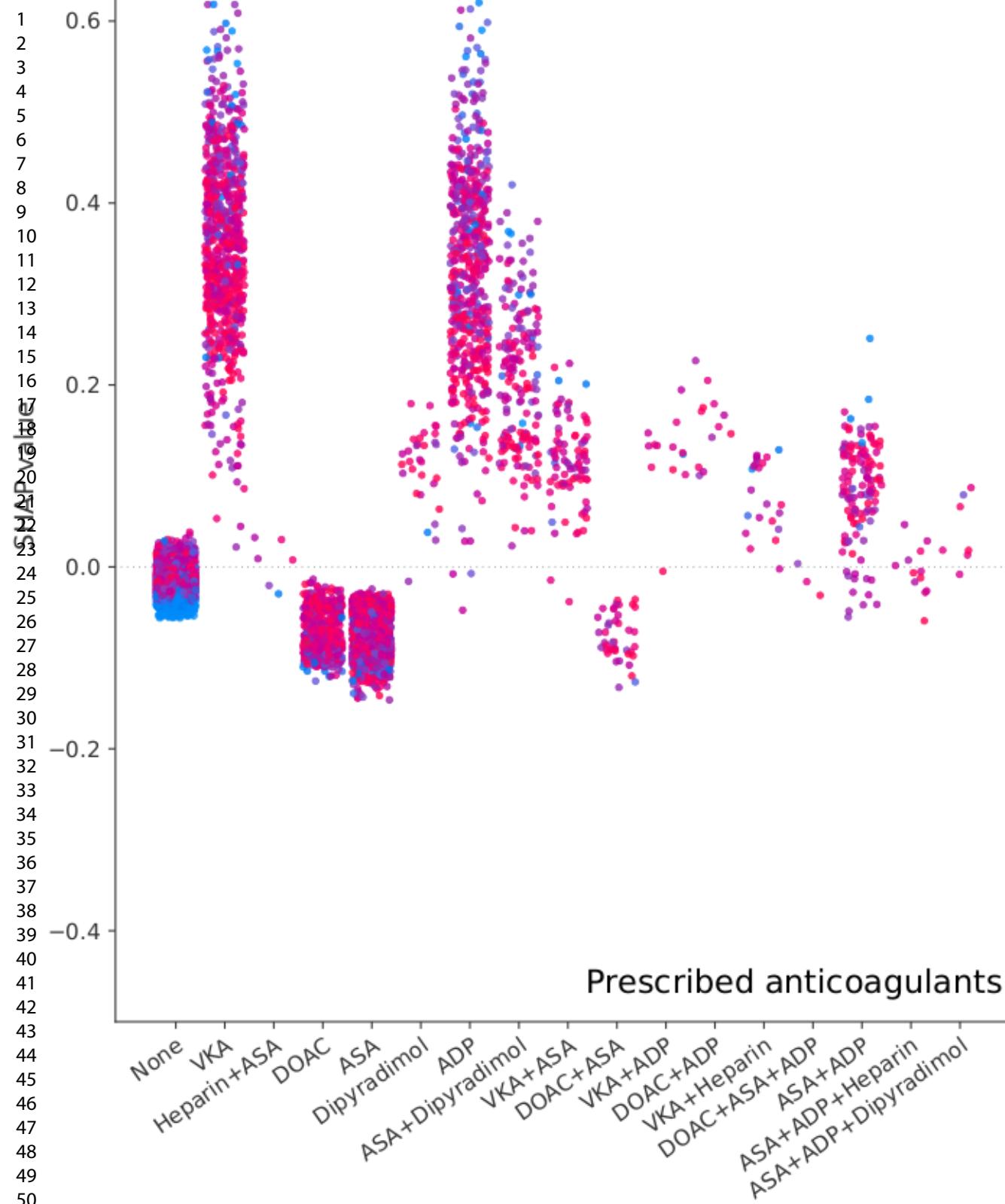
297x209mm (300 x 300 DPI)



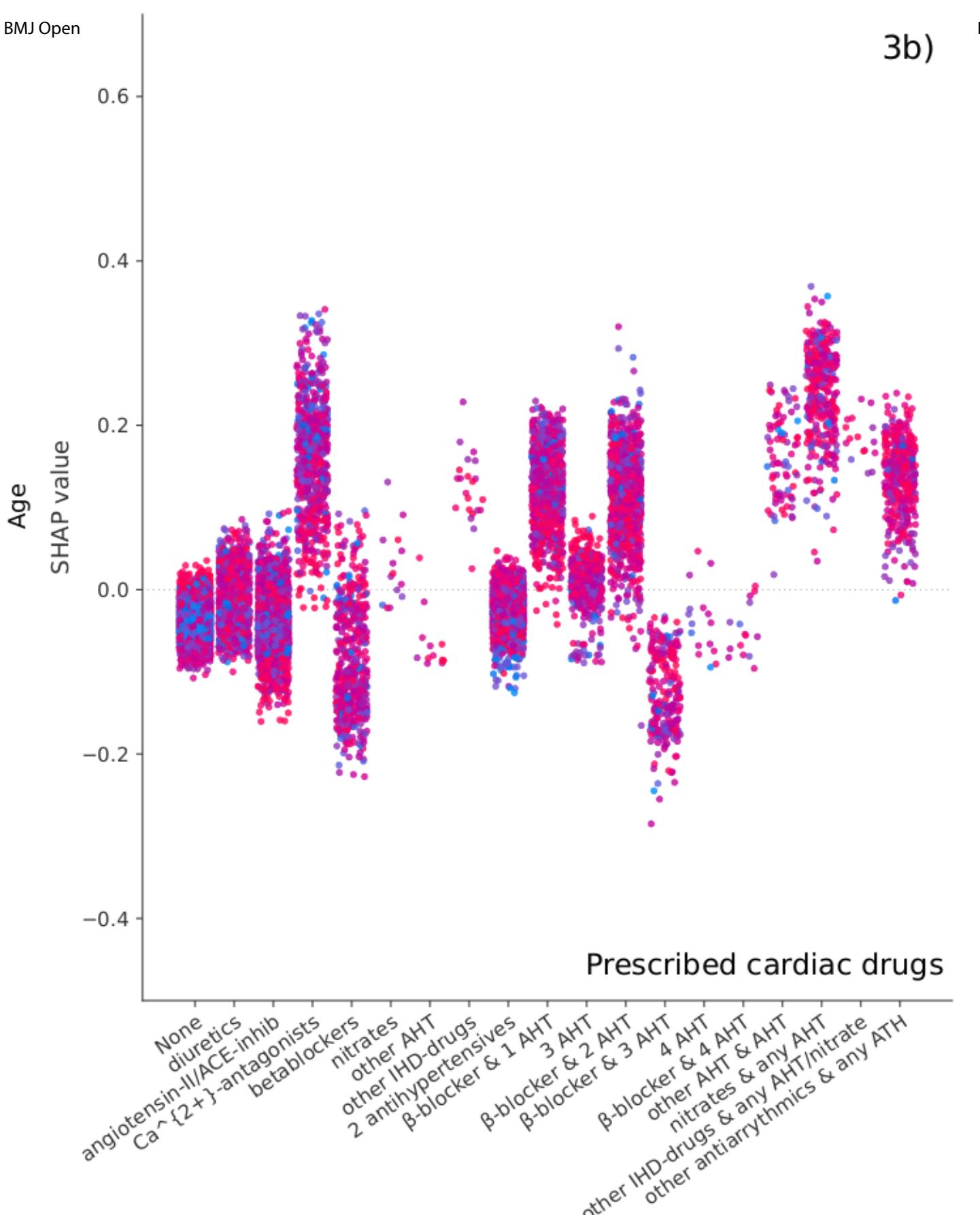
2a) The overall importance of the 10 most important variables measured by the SHAP-values for the full machine-learning and full logistic regression models on the primary outcome (LOS >4 days or readmission due to "medical" morbidity). Only the importance of prescribed anticholesterols and gender differ between the models. The contributions of the remaining variables are summed in the bottom bar.
 2b) The SHAP-values for the full machine-learning model on the primary outcome, where positive increase and negative values decrease the risk score. Each dot represents a patient and the color is related to the value of the variable with blue being lowest and red highest.

297x209mm (300 x 300 DPI)

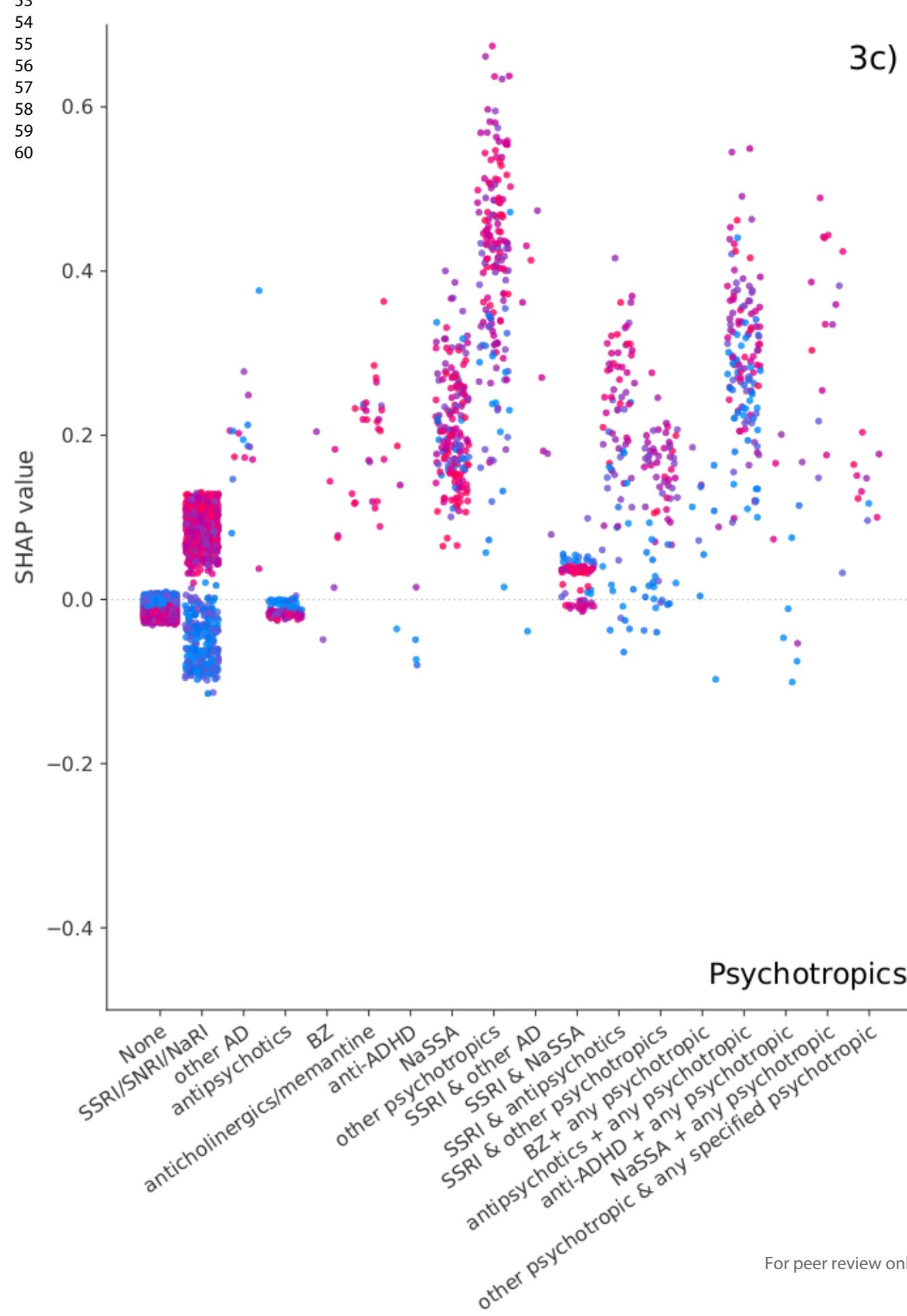
3a)



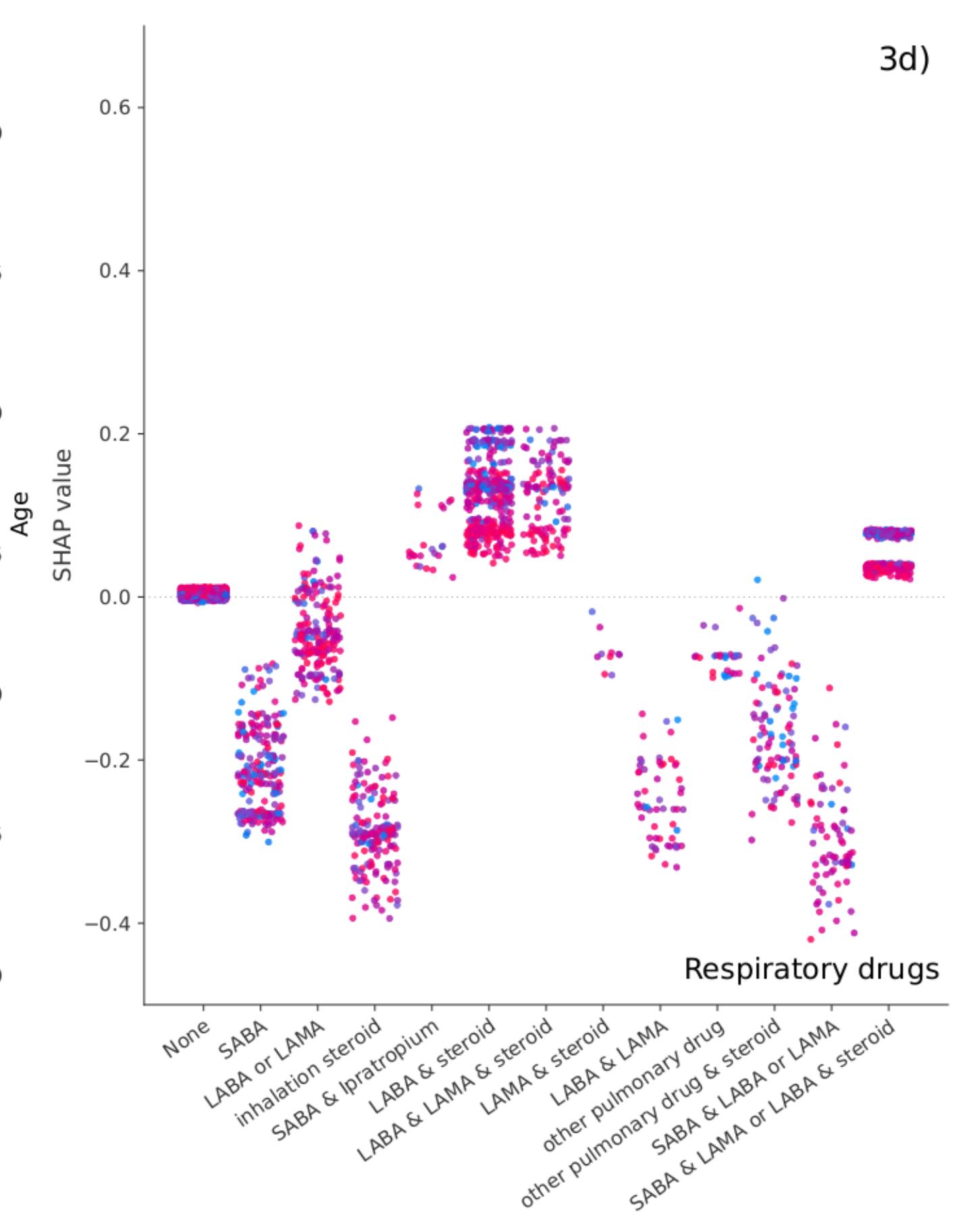
3b)

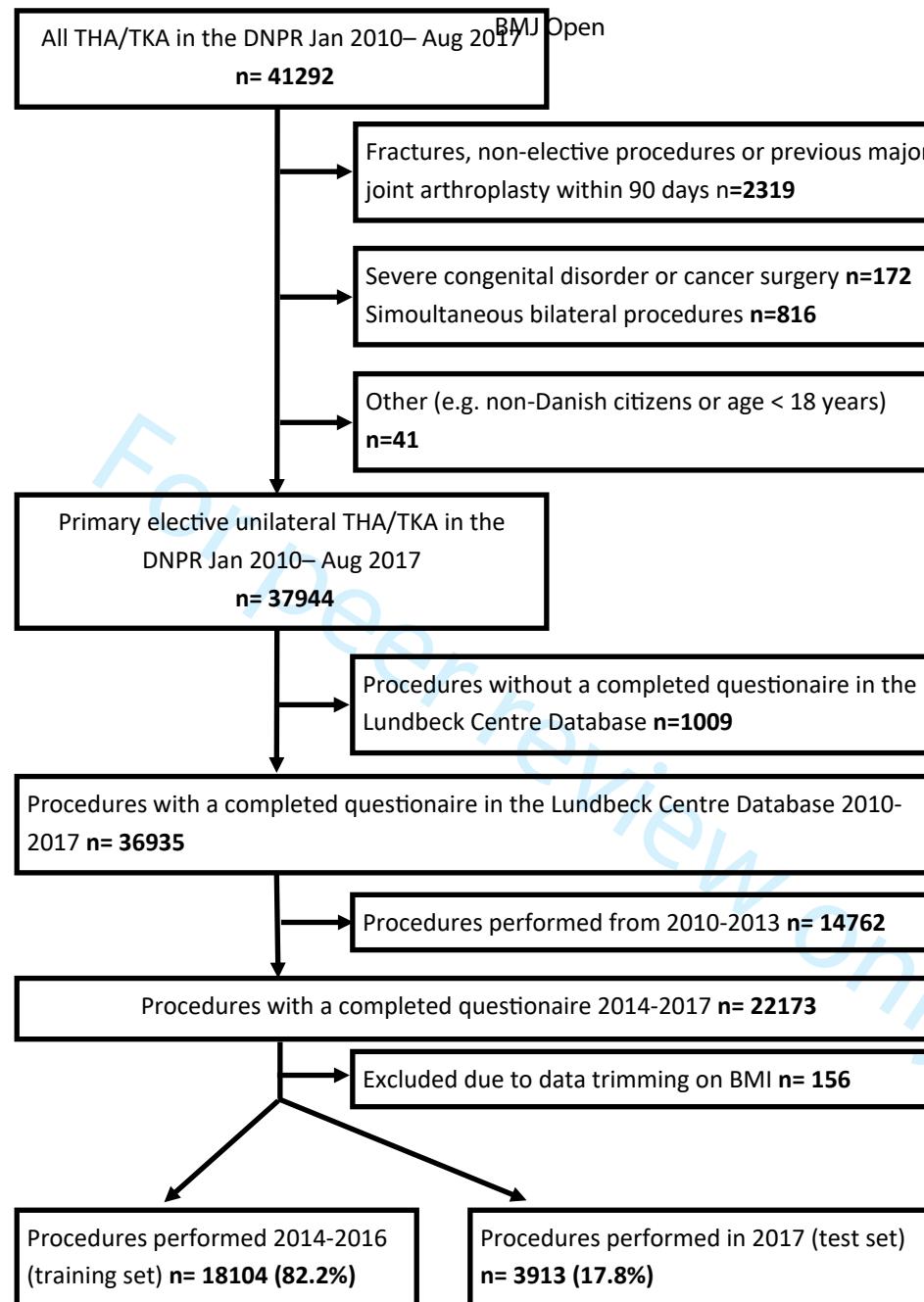


3c)

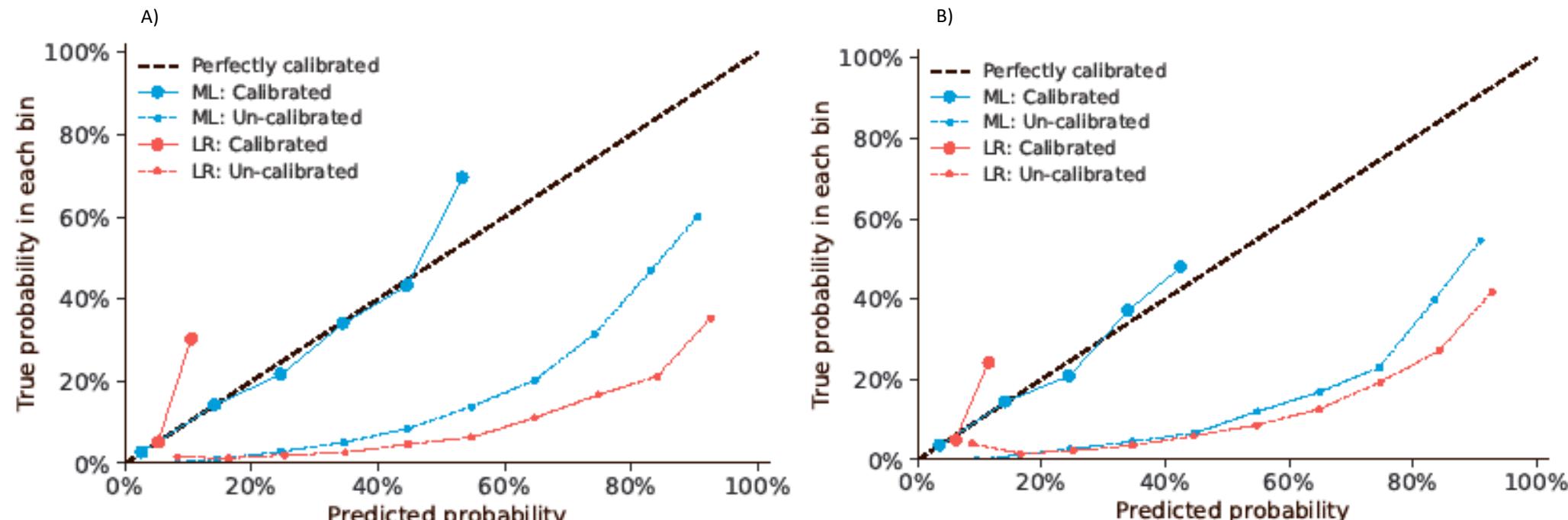


3d)





Flowchart of the study population and final sample size. THA: total hip arthroplasty TKA: total knee arthroplasty DNPR: the Danish National Patient Registry
For peer review only <http://bmjopen.bmjjournals.com/site/about/guidelines.xhtml>

1
2 Supplemental Material 2
3
4
5

Calibration plot of the machine learning model (ML) and the logistic regression (LR) for both the calibrated and un-calibrated models for the primary (A) and secondary (B) outcome.

1
2
3
4 **Supplemental material table 3**
5

6 Details on specific drugs with reimbursed prescriptions 6 months preoperatively.
7 Numbers are n (%)

8 Reimbursed prescriptions within 3 months preoperatively	9	10 training set (n:18104)	11 test set (n:3913)
Anticoagulants			
none		13570 (75.0)	2953 (75.5)
VKA		729 (4.0)	127 (3.2)
Heparin & Acetylsalicylic acid		6 (0.0)	1 (0.0)
DOAC		526 (2.9)	181 (4.6)
Acetylsalicylic acid		2235 (12.3)	462 (11.8)
Dipyradimol		31 (0.2)	3 (0.1)
ADP-antagonist		522 (2.9)	122 (3.1)
Acetylsalicylic acid & Dipyradimol		169 (0.9)	16 (0.4)
VKA & Acetylsalicylic acid		81 (0.4)	7 (0.2)
DOAC & Acetylsalicylic acid		42 (0.2)	5 (0.1)
VKA & ADP-antagonist		13 (0.1)	2 (0.1)
DOAC & ADP-antagonist		12 (0.1)	5 (0.1)
VKA & Heparin		22 (0.1)	0 (0.0)
DOAC & Acetylsalicylic acid & ADP-antagonist		3 (0.0)	1 (0.0)
Acetylsalicylic acid & ADP-antagonist		124 (0.7)	26 (0.7)
Acetylsalicylic acid & ADP-antagonist & Heparin		12 (0.1)	1 (0.0)
Acetylsalicylic acid & ADP-antagonist & Dipyradimol		7 (0.0)	1 (0.0)
Cardiac prescriptions			
none		7741 (42.8)	1780 (45.5)
diuretics		1070 (5.9)	191 (4.9)
angiotensin-II/ACE-inhibitors		2287 (12.6)	528 (13.5)
Ca ²⁺ antagonists		688 (3.8)	140 (3.6)
β-blocker		492 (2.7)	96 (2.5)
nitrates		14 (0.1)	5 (0.1)
other antihypertensives		12 (0.1)	0 (0.0)
other types of medication for IHD		22 (0.1)	1 (0.0)
2 antihypertensives		2360 (13.0)	513 (13.1)
β-blocker & 1 antihypertensive ¹		966 (5.3)	195 (5.0)
3 antihypertensives		515 (2.8)	83 (2.1)
β-blocker & 2 antihypertensives ¹		902 (5.0)	168 (4.3)
β-blocker & 3 antihypertensives ¹		235 (1.3)	55 (1.4)
4 antihypertensives		16 (0.1)	4 (0.1)
β-blocker & 4 antihypertensives		15 (0.1)	6 (0.2)
other antihypertensive & antihypertensives ¹		78 (0.4)	18 (0.5)
nitrates & any hypertensive		323 (1.8)	57 (1.5)
other drugs for IHD & any antihypertensive and/or nitrate		16 (0.1)	4 (0.1)
other antiarrhythmics & any antihypertensives		352 (1.9)	69 (1.8)
Anticholesterols			
none		12665 (70.0)	2762 (70.6)
statins		5218 (28.8)	1105 (28.2)
other anti-lipids		118 (0.7)	24 (0.6)
statins +other anti-lipids		103 (0.6)	22 (0.6)
Systemic steroids			
		1038 (5.7)	234 (6.0)

	<u>Antirheumatics</u>		
5	none	17709 (97.8)	3822 (97.7)
6	disease-modifying antirheumatic drugs	392 (2.2)	91 (2.3)
7	other antirheumatics	3 (0.0)	0 (0.0)
9	<u>Respiratory prescriptions</u>		
10	none	16256 (89.8)	3498 (89.4)
11	SABA	235 (1.3)	54 (1.4)
12	LABA or LAMA	194 (1.1)	42 (1.1)
13	inhalation steroid only	176 (1.0)	43 (1.1)
14	SABA & Ipratropium (+/- others)	24 (0.1)	0 (0.0)
15	LABA & steroid	432 (2.4)	87 (2.2)
16	LABA & LAMA & steroid	115 (0.6)	26 (0.7)
17	LAMA & steroid	10 (0.1)	1 (0.0)
18	LABA & LAMA	56 (0.3)	31 (0.8)
19	other pulmonary drugs	26 (0.1)	9 (0.2)
20	other pulmonary drugs & steroid	95 (0.5)	12 (0.3)
21	SABA & LABA or LAMA	76 (0.4)	26 (0.7)
22	SABA & LABA or LAMA & steroid	409 (2.3)	84 (2.1)
24	<u>Psychotropic prescriptions</u>		
25	none	16113 (89.0)	3496 (89.3)
26	SSRI/SNRI/NaRI	1055 (5.8)	209 (5.3)
27	other antidepressants	16 (0.1)	2 (0.1)
28	antipsychotics	104 (0.6)	20 (0.5)
29	benzodiazepines ²	7 (0.0)	0 (0.0)
30	anti-cholinergics or memantine	27 (0.1)	6 (0.2)
31	anti-ADHD drugs	7 (0.0)	4 (0.1)
32	NaSSA	177 (1.0)	32 (0.8)
33	other psychotropics	166 (0.9)	44 (1.1)
34	SSRI + other antidepressants	9 (0.0)	1 (0.0)
35	SSRI + NaSSA	86 (0.5)	16 (0.4)
36	SRRI + antipsychotics	80 (0.4)	18 (0.5)
37	SRRI + other psychotropics	72 (0.4)	19 (0.5)
38	benzodiazepines + any psychotropic	11 (0.1)	4 (0.1)
39	antipsychotics + any psychotropic	137 (0.8)	32 (0.8)
40	anti-ADHD + any psychotropic	11 (0.1)	3 (0.1)
41	NaSSA + any psychotropic	16 (0.1)	6 (0.2)
42	other psychotropics + any specified psychotropic	10 (0.1)	1 (0.0)

VKA: vitamin K antagonists DOAC: direct oral anticoagulant ADP: Adenosine diphosphate 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 ²likely underreported due to limited general reimbursement for benzodiazepines in Denmark

1
2
3
4 **Supplemental material 4**
5

6 Performance of the different models with a predefined positive prediction fraction of 25 and 30 for the primary outcome (LOS >4 days or
7 readmission due to "medical" morbidity.

Positive prediction fraction 25%	TP	FP	FN	TN	Sensitivity %	Precision %	MCC %	AUROC %	AUPRC %	Brier %	P(sensitivity) %
Full machine-learning model	120	858	62	2873	65.9	12.3	20.9	77.0	15.3	4.32	-
Full logistic regression model	108	870	74	2861	59.3	11.0	17.5	74.6	15.6	4.32	10.4
Parsimonious machine-learning model	114	864	68	2867	62.6	11.7	19.2	74.9	14.1	4.35	26.2
Parsimonious logistic regression model	103	875	79	2856	56.6	10.5	16.1	73.6	15.2	4.33	3.9
Age-model	94	824	88	2907	51.6	10.2	14.7	69.7	12.2	38.8	1.2

Positive prediction fraction 30%	TP	FP	FN	TN	Sensitivity %	Precision %	MCC %	AUROC %	AUPRC %	Brier %	P(sensitivity) %
Full machine-learning model	130	1043	52	2688	71.4	11.1	20.0	77.0	15.3	4.32	-
Full logistic regression model	117	1056	65	2675	64.2	10.0	16.5	74.6	15.6	4.32	8.4
Parsimonious machine-learning model	124	1049	58	2682	68.1	10.6	18.4	74.9	14.1	4.35	30.0
Parsimonious logistic regression model	118	1055	64	2676	64.8	10.1	16.8	73.6	15.2	4.33	14.0
Age-model	100	955	82	2776	54.9	9.5	13.9	69.7	12.2	38.8	0.9

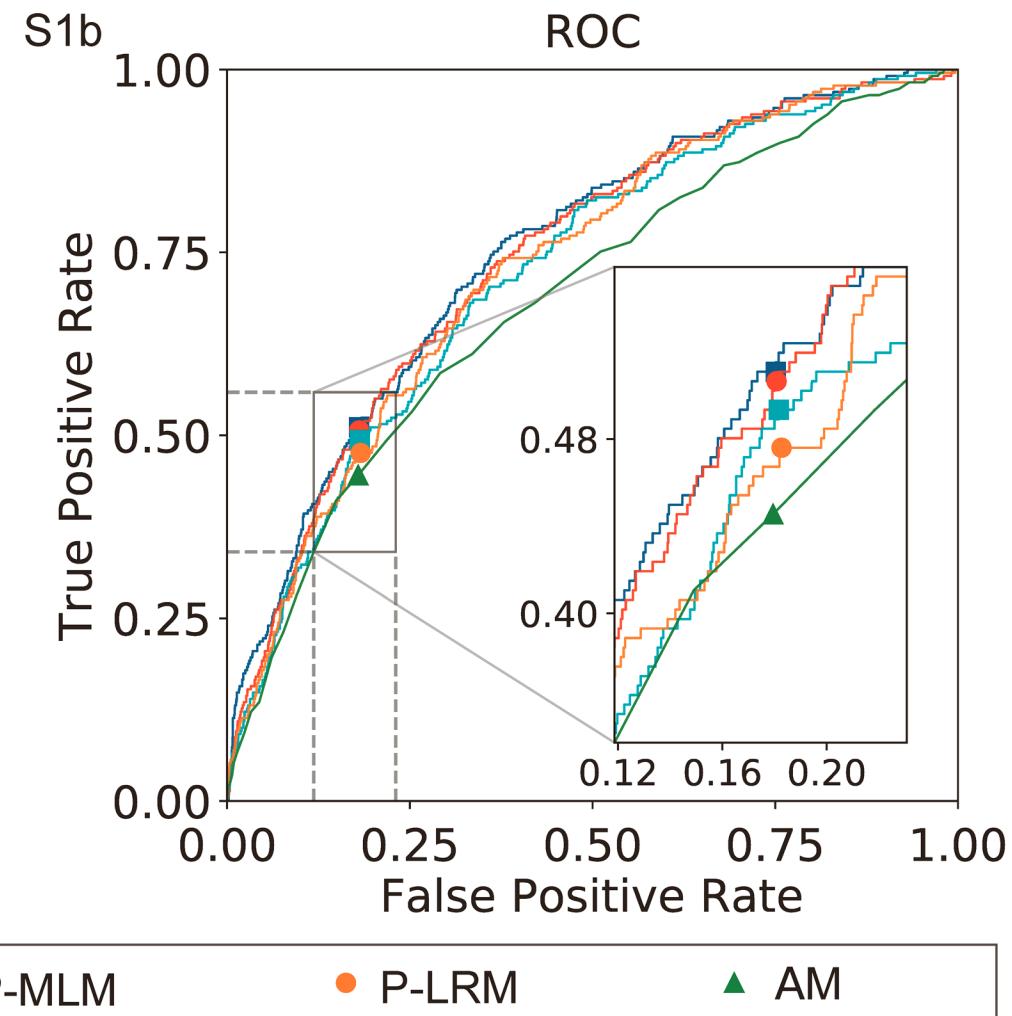
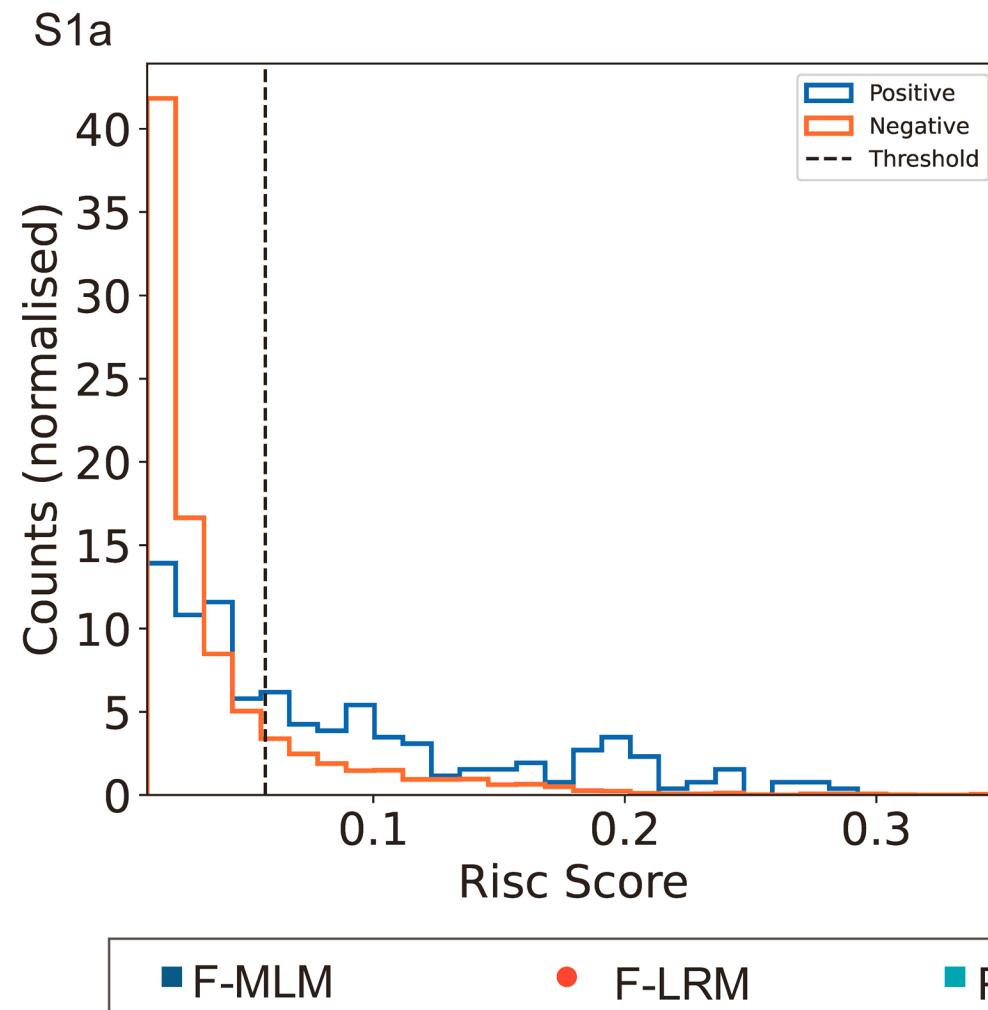
35 TP: true positives FP: false positives FN: false negatives TN: true negatives MCC: Matthews correlation coefficient AUROC: area under the
36 receiver operating curve AUPRC: area under the precision recall curve P(sensitivity): probability that the model performs better than the
37 machine-learning model relative to sensitivity.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supplemental material 5

Table S1: Performance of different models for the secondary outcome (LOS >4 days or readmissions due to "medical" morbidity or LOS >4 days but without recorded morbidity)

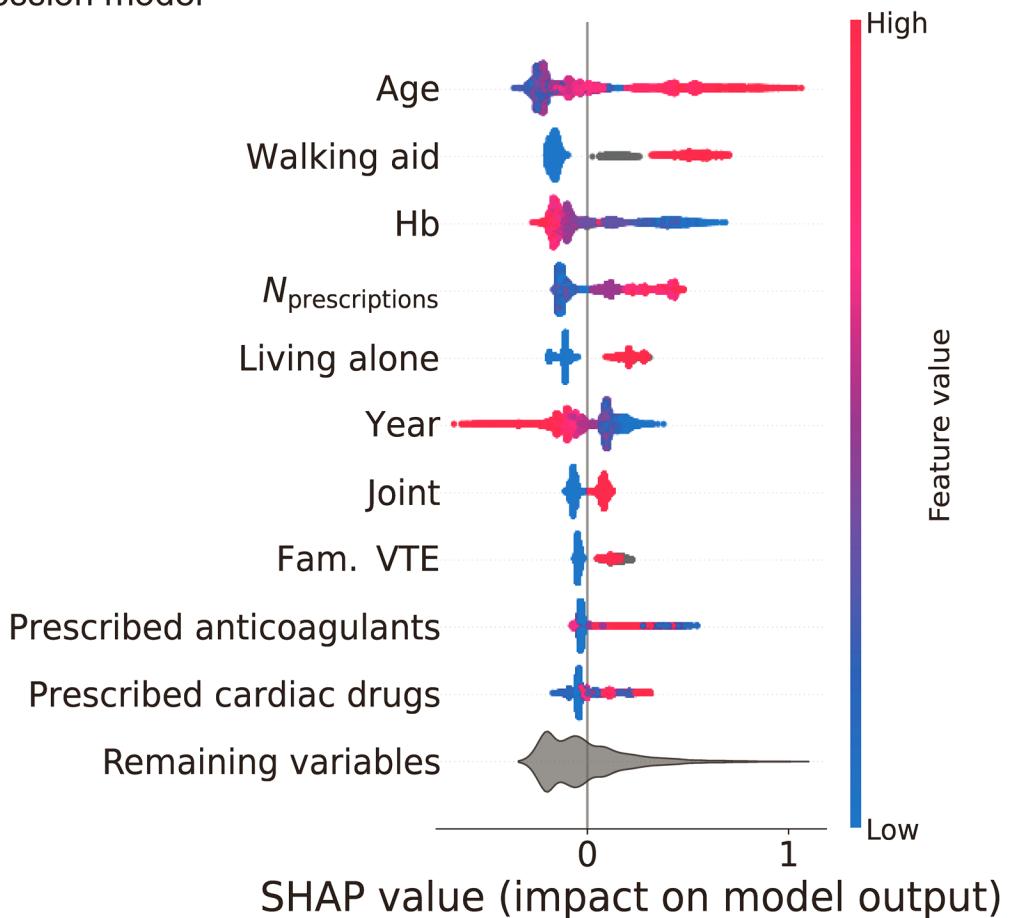
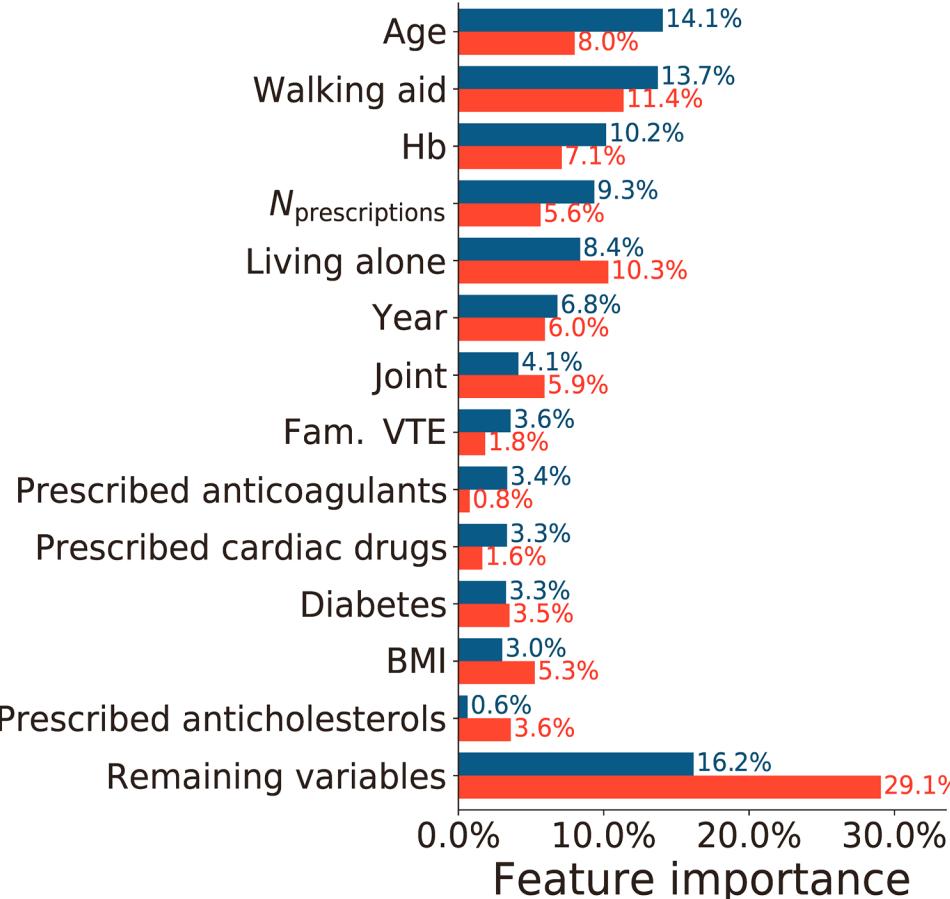
Positive prediction fraction 20%	TP	FP	FN	TN	Sensitivity %	Precision %	MCC %	AUROC %	AUPRC %	Brier %	P(sensitivity) %
Full machine-learning model	117	665	112	3019	51.1	15.0	19.4	75.0	18.1	5.23	-
Full logistic regression model	115	667	114	3017	50.2	14.7	18.9	74.1	16.7	5.35	46.4
Parsimonious machine-learning model	109	673	120	3011	47.6	13.9	17.2	72.1	15.8	5.33	35.2
Parsimonious logistic regression model	109	673	120	3011	47.6	13.9	17.2	72.9	16.7	5.37	22.6
Age-model	102	661	127	3023	44.5	13.4	15.8	68.7	13.4	38.3	10.3
Positive prediction fraction 25%	TP	FP	FN	TN	Sensitivity %	Precision %	MCC %	AUROC %	AUPRC %	Brier %	P(sensitivity) %
Full machine-learning model	128	850	101	2834	55.9	13.1	17.8	75.0	18.1	5.23	-
Full logistic regression model	133	845	96	2839	58.1	13.6	19.1	74.1	16.7	5.35	68.0
Parsimonious machine-learning model	121	857	108	2827	52.8	12.3	16.3	72.1	15.8	5.33	25.5
Parsimonious logistic regression model	127	851	102	2833	55.5	13.0	17.5	72.9	16.7	5.37	46.6
Age-model	113	805	116	2879	49.3	12.3	15.2	68.7	13.4	38.3	17.2
Positive prediction fraction 30%	TP	FP	FN	TN	Sensitivity %	Precision %	MCC %	AUROC %	AUPRC %	Brier %	P(sensitivity) %
Full machine-learning model	146	1027	83	2657	63.4	12.4	18.4	75.0	18.1	5.23	-
Full logistic regression model	144	1029	85	2655	62.9	12.3	17.9	74.1	16.7	5.35	42.4
Parsimonious machine-learning model	135	1038	94	2651	59.0	11.5	15.8	72.1	15.8	5.33	14.9
Parsimonious logistic regression model	140	1033	89	2651	61.1	11.9	17.0	72.9	16.7	5.37	28.3
Age-model	122	933	107	2751	53.3	11.6	14.8	68.7	13.4	38.3	7.9

TP: true positives FP: false positives FN: false negatives TN: true negatives MCC: Matthews correlation coefficient AUROC: area under the receiver operating curve AUPRC: area under the precision recall curve P(sensitivity): probability that a model performs better than the machine-learning model relative to sensitivity.

1
2
3
4 Supplemental material 6 Figure S1a-b
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

S1a) Distribution of full machine-learning model risk-scores for patients +/- the secondary outcome (LOS >4 days or readmissions due to "medical" morbidity or LOS >4 days with no recorded morbidity). The dashed line marks the classification threshold of a 20% positive prediction fraction.

S1b) Receiver operating curves (ROC) for the full machine-learning model (F-MLM), full logistic regression model (F-LRM), parsimonious machine-learning model (P-MLM), parsimonious logistic regression model (P-LRM) and the age-only model (AM).

1
2 Supplemental Material 7
34 Figure S2a-b
56 S2a Full Machine-learning model Full logistic regression model S2b
7

S2a) The overall importance of the 10 most important variables measured by the SHAP-values for the full machine-learning and full logistic regression models for the secondary outcome (LOS > 4 days or readmissions due to "medical" morbidity or LOS > 4 days with no recorded morbidity).

Only the importance of prescribed anti-cholesterols and familiar disposition for venous thromboembolism differed between the models. The contributions of the remaining variables are summed in the bottom bar.

S2b) The SHAP-values for the full machine-learning model. Positive SHAP-values increase the risk score while negative values decrease the risk score. Each dot repre-

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supplemental material 8

Figure S3a-d

SHAP scatter-plot on the contributions to the full machine-learning model on outcome B (LOS >4 days or readmission due to “medical” morbidity), for individual types of prescribed anticoagulants, cardiac drugs, psychotropics and respiratory drugs stratified by age.

Legend:

3a) Prescribed anticoagulants

VKA: vitamin K antagonists ASA: acetylsalicylic acid DOAC: direct oral anticoagulant ADP: Adenosine diphosphate ACE: angiotensin converting enzyme

3b) Prescribed cardiac drugs

ACE: angiotensin converting enzyme AHT: antihypertensive. Other AHT were defined as AHT different from diuretics ANG-II/ACE inhibitors or Ca²⁺-antagonists. IHD: Ischemic heart disease

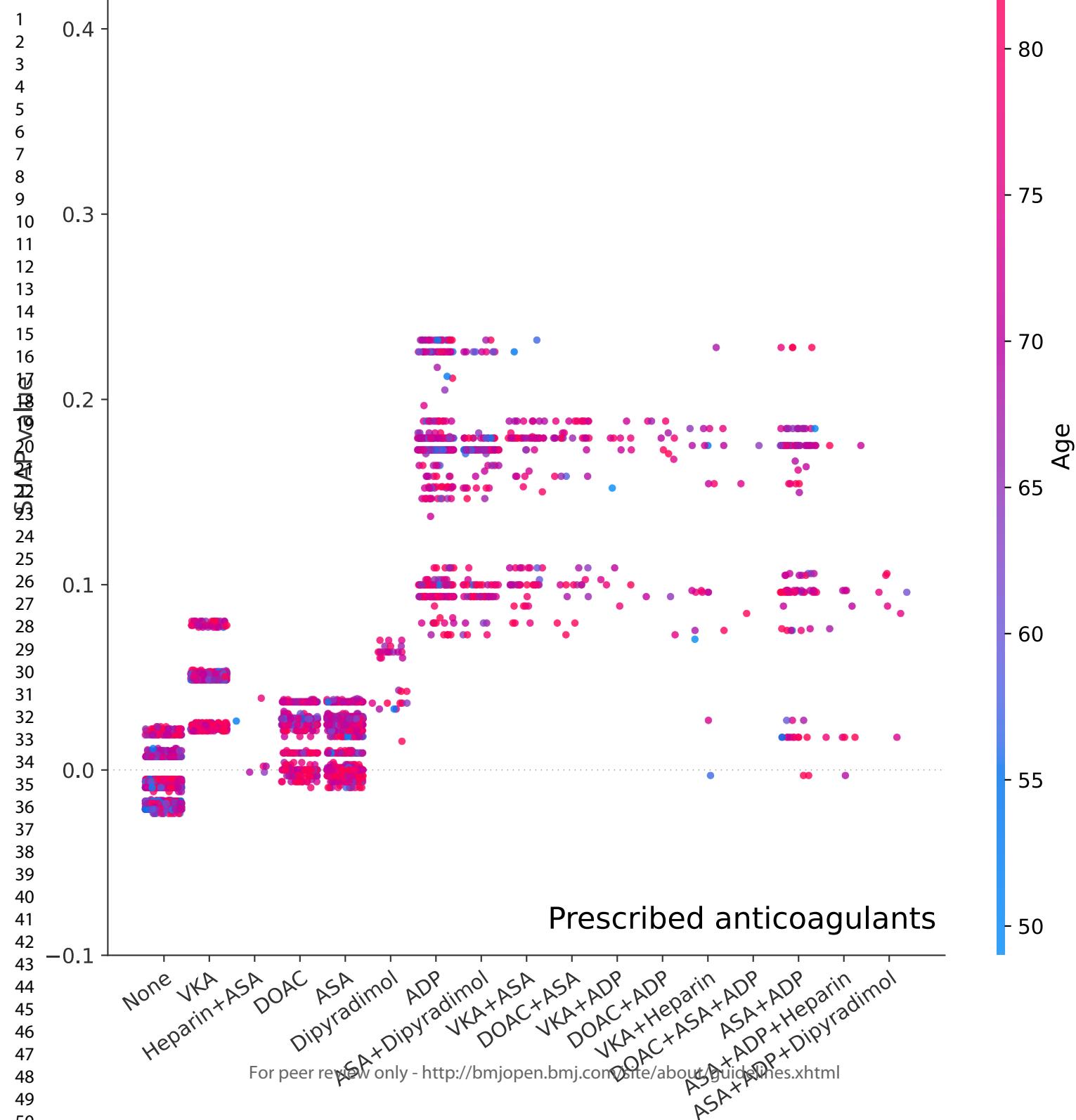
3c) Prescribed psychotropics

SSRI: Selective serotonin inhibitor SNRI: Serotonin and norepinephrine reuptake inhibitor NaRI: Norepinephrine reuptake inhibitor NaSSA: Norepinephrine and specific serotonergic antidepressants. AD: antidepressants BZ: Benzodiazepines (likely underreported due to limited general reimbursement in Denmark). ADHD: Attention-deficit/hyperactivity disorder

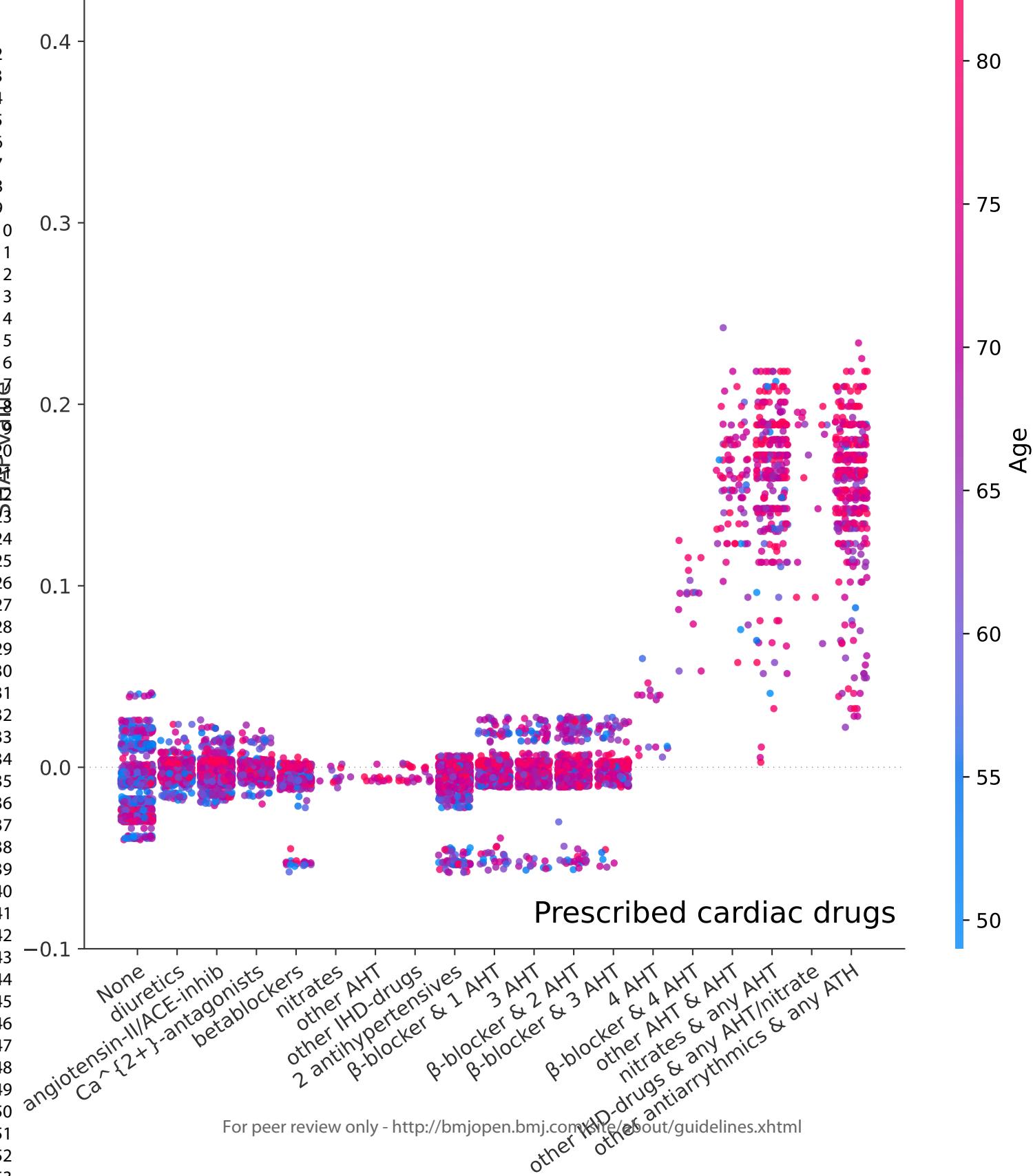
3d) Prescribed respiratory drugs

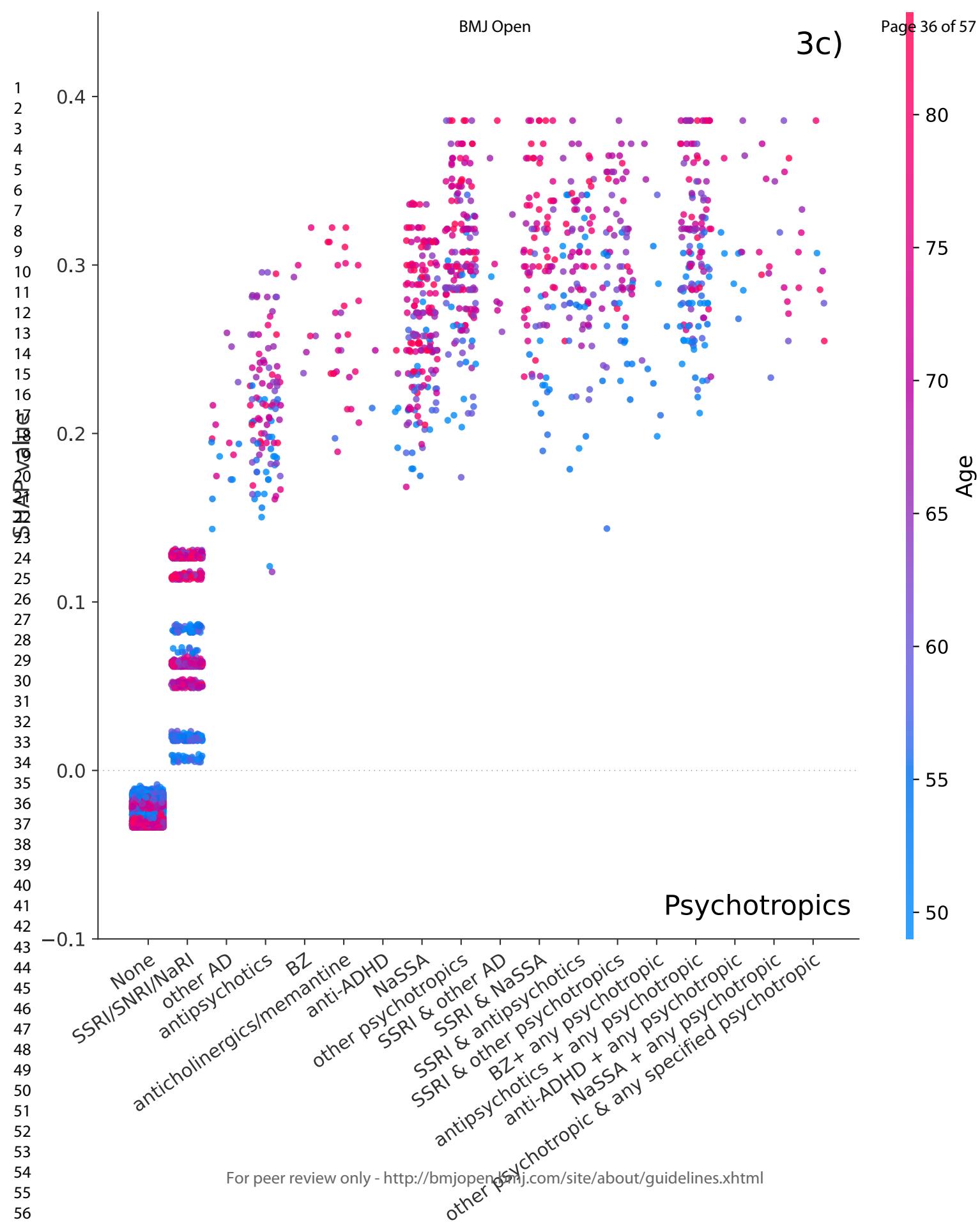
The model found no additional information from this variable why all values equal 0.

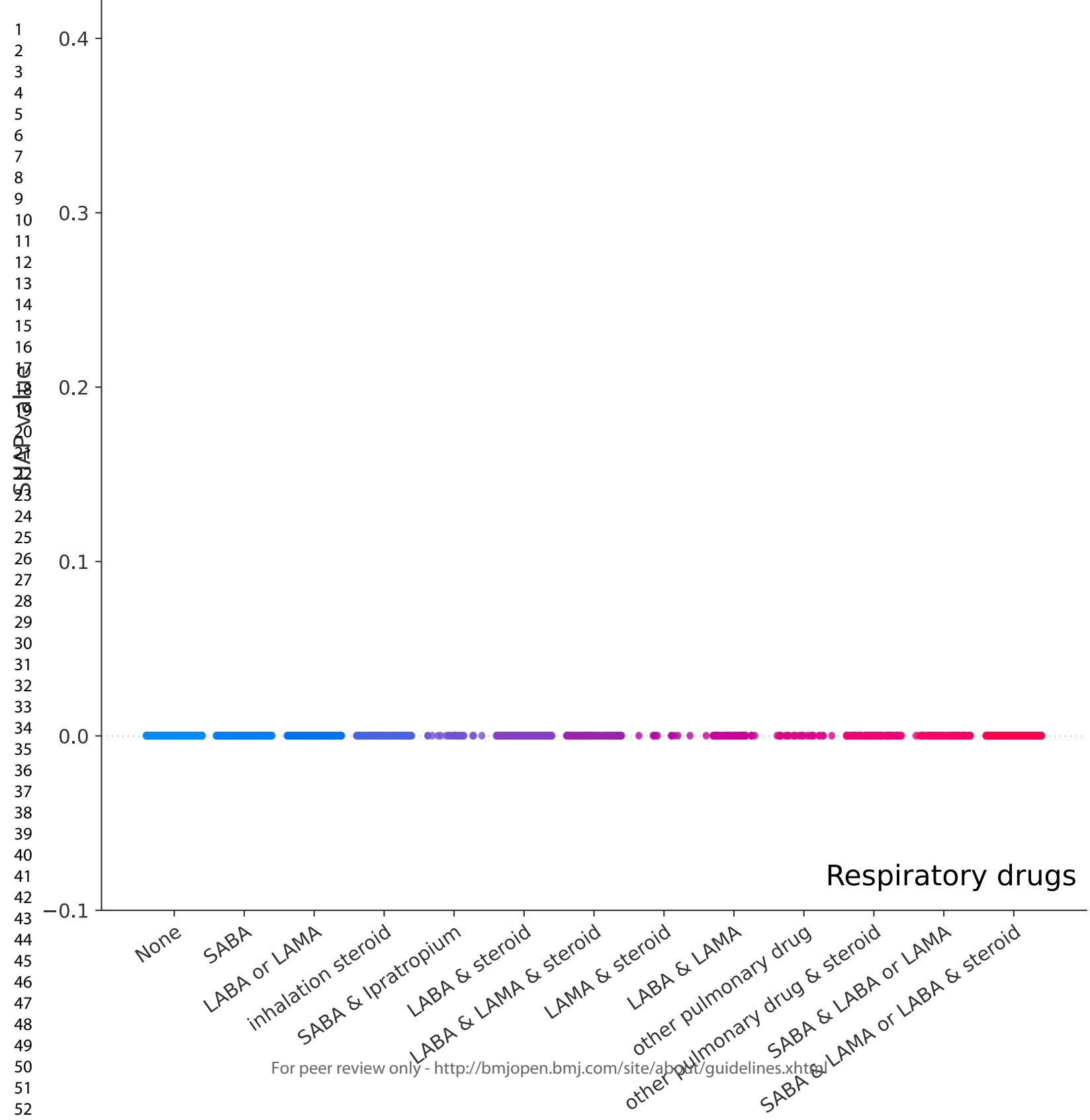
SABA: Short-acting beta agonist LABA: long-acting beta agonist LAMA: Long-acting muscarinic antagonist.



3b)







TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item	Checklist Item		Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	4-5
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	4
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	5
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	5
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	5
	5b	D;V	Describe eligibility criteria for participants.	5
	5c	D;V	Give details of treatments received, if relevant.	5
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	6
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	N/A
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	7
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	D;V	Explain how the study size was arrived at.	N/A
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	7
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	7
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	7
	10c	V	For validation, describe how the predictions were calculated.	7
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	8
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	7 and Sup mat 2
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	Tbl1
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	Sup mat 1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	Tbl 1
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	Tbl 1
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	7-9
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	7
	15b	D	Explain how to use the prediction model.	Fig 2a-b
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	Tbl2
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N/A
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	14
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	10-14
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	14
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	16
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	16

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.

For peer review only - <http://bmjopen.bmjjournals.org/site/about/guidelines.xhtml>