Anesthesiology

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

--Manuscript Draft--

Manuscript Number:	
Full Title:	Preoperative prediction of medical morbidity after fast-track hip and knee arthroplasty - a machine learning based approach.
Short Title:	Machine-learning models in joint replacement
Article Type:	Original Investigation: Perioperative Medicine
Section/Category:	
Corresponding Author:	Christoffer Calov Jorgensen, M.D. Rigshospitalet Copenhagen, DENMARK
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Rigshospitalet
Corresponding Author's Secondary Institution:	
First Author:	Christian Michelsen, M.Sci
First Author Secondary Information:	
Order of Authors:	Christian Michelsen, M.Sci
	Christoffer Calov Jorgensen, M.D.
	Mathias Heltberg, M.Sci
	Mogens H. Jensen, D.Sci
	Alessandra Lucchetti, M.Sci
	Pelle Baggesgaard Petersen, M.D., Ph.D
	Troels Christian Petersen, M.Sci
	Henrik Kehlet, M.D., Ph.D
Order of Authors Secondary Information:	
Suggested Reviewers:	Stavros Memtsoudis, M.D. Director of Critical Care Services, HSS: Hospital for Special Surgery memtsoudiss@hss.edu Dr. Memtsoudis is an international expert on perioperative care in hip and knee arthroplasty.
	Asokumar Buvanendran, M.D. Rush Medical College of Rush University: Rush University Rush Medical College asokumar@aol.com
	Lee Fleisher, M.D. University of Pennsylvania Perelman School of Medicine Lee.Fleisher@pennmedicine.upenn.edu
Opposed Reviewers:	

Dear Dr. Kharasch

Enclosed is our manuscript entitled:" Preoperative prediction of medical morbidity after fast-track hip and knee arthroplasty - a machine learning based approach." This study investigates potential advantages of a state of the art machine-learning model comprising 33 preoperative variables and including novel use of preoperative dispensed prescriptions within 3 months preoperatively, for prediction of postoperative medical complications resulting in prolonged hospitalizations or readmissions in fully implemented fast-track total hip and knee replacement. We believe that our results are an important contribution within the field of perioperative medicine and risk-prediction, especially the novel analyzes on the specific contributions of individual risk-factors in the machine-learning model. Consequently, we hope you will consider our study for publication in Anesthesiology.

We are aware of the large number of figures, most of which are Supplemental Digital Content. However, due to difficulties in combining the 4 panels of Figure 3a-d into a single PDF-file these have been submitted as separate files. We hope that you are able to assist in combining these panels into a single figure during the editorial process in case of acceptance.

Kind regards, on behalf of the authors

Christoffer Jørgensen and Henrik Kehlet

Manuscript (Title Page, Abstract, Body, References, Figure Legends)

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

Pelle B Petersen, 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, M.D., Ph.D., 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 Orthopedic Surgery, Lillebaelt Hospital - Vejle, University Hospital of Southern Denmark, 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, Sydvestjysk Hospital, Grindsted, Denmark

7. Word and Element Counts:

Abstract: 300/300 Introduction: 466/500 Discussion:1278/1500 Figures:3 Tables:2 Appendices:2 Supplementary Digital Files:4

8. Abbreviated title: Machine learning models in joint replacement

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. Dr. Petersen is an advisory member of Sanofi outside of the present study.

Abstract

Background: Introduction of machine-learning models has potentially improved prediction of postoperative hospitalization and morbidity after hip and knee replacement. 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 machine-learning models for prediction of postoperative morbidity in enhanced recovery hip and knee replacement remain uncertain.

Methods: Multicenter cohort study from 2014-2017 in enhanced recovery total hip and knee replacement. Prospective recording of comorbidity and prescriptions. Information on length of stay and readmissions through the Danish National Patient Registry and medical records. Data was split into training (n:18013) and test sets (n:3913). A machine-learning model with 33 variables was used for predicting “medical” morbidity with a length of stay of >4 days or 90-days readmission and compared to a full logistic regression model. In addition, a machine-learning model excluding age, an age-only model and parsimonious machine-learning and logistic regression models using the ten most important variables were evaluated. Model performances were evaluated using several metrics, including precision, operating receiver (AUC) and precision recall curves (AUPRC). Variable importance was analyzed using Shapley Additive Explanations values.

Results: With 782 (20%) “risk-patients”, precision, AUC and AUPRC were 13.6%, 76.3% and 15.5% for the full and 12,8%, 75.9% and 17.1% for the parsimonious machine-learning models vs. 12.5%, 74.5% and 15.7% for the full logistic regression model. The machine-learning model excluding age and the Age-only model performed worse. Of the top ten variables, eight were shared between the full machine-learning and logistic regression models, and the importance of specific prescribed drugs varied considerably with age.

Conclusion: A machine-learning algorithm using preoperative characteristics and prescriptions likely improves identification of patients in high-risk of medical complications after fast-track hip and knee replacement. Such algorithms could help identify patients who benefit 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, 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 total hip replacement, total knee replacement 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] readmissions[16] or payment models,[17,18] have shown promising predictive value compared to conventional statistical methods.[19]

However, few 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. In our previous study of 9512 total hip and knee replacements within an enhanced recovery 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 analyzed whether an improved machine-learning model was better for preoperative prediction of medical complications resulting in prolonged length of stay and readmissions compared to a traditional logistic regression model, in a large consecutive cohort of patients undergoing fast-track total hip and knee replacement within a national public health-care system.[1] In addition to well-defined patient-reported preoperative risk-factors, we also included information on dispensed reimbursed prescriptions 6 months prior to surgery using a nationwide registry.[21]

Method

Reporting of the study is done in accordance with the Transparent reporting of multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement[22] and the Clinical AI Research (CAIR) checklist proposal.[23]

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 a preoperative questionnaire with nurse assistance if needed. 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.[21] Finally, data were combined with the Danish National Patient Registry (DNPR) for information on length of stay (counted as postoperative nights spent in hospital), 90-days readmissions with overnight stay and mortality. In case of length of stay >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 reviewed to identify cause of death. Evaluation of discharge and medical records was performed by PP supervised by CJ. In case of disagreement, records were conferred with HK. Subsequently, causes of length of stay >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 length of stay 4-6 days with a standard discharge summary describing a successful postoperative course, it was assumed that no clinically relevant postoperative complications had occurred. If length of stay was >6 days but with standard discharge summary, the entire medical record was evaluated to confirm that no relevant 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 total hip and knee replacement 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 length of stay ≤5 days, early mobilization, functional discharge criteria and discharge to own home.[1] There was 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 total hip or total knee replacement and total hip replacement due to severe congenital joint disorder or cancer.

Outcomes

The primary outcome was to compare prediction quality when using a machine-learning model to predict the occurrence of “medical” complications resulting in a length of stay >4 days or readmission compared to a traditional logistic regression model (outcome A). Secondarily, we investigated how inclusion of cases with a length of stay >4 days but no reported “medical” complication as a positive outcome influenced the model (outcome B). For both outcomes, we also investigated whether a parsimonious model including only the top ten variables would perform equally well as the full model, and whether the effect of age per se would compare to the full machine-learning model. All figures and tables in the main text and Appendix are based on outcome A; the corresponding figures for outcome B are reported in the Supplemental Digital Content.

Statistical Analysis

Data was initially trimmed 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, the dataset was subsequently split into a training set consisting of 18.013 (82.2%) procedures from 2014-2016 and a test set of 3913 (17.8%) procedures from 2017.

As a reference model, we used classical logistic regression using all 33 input variables (table 1). Cases of missing values in the logistic regression model 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)[24] for the machine-learning models, as such methods work well with categorical data and missing values. We tried using both normal cross entropy and FocalLoss[25] as the objective function for the machine-learning model. The reason for testing FocalLoss was to allow the machine-learning model to focus more on the (few) positives.

The full machine-learning model was trained and hyperparameter optimized using state of the art machine-learning 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 were predicted as positive by the model (a positive predictive fraction (PPF) of 20%). We also included results for PPF values of 25% and 30%. Furthermore, we trained two parsimonious models using machine-learning and logistic regression with only the 10 most important features. Finally, we specifically explored the influence of increasing age, by constructing a model based only on age (Age), and a machine-learning model based on all variables except for age.

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.[26,27] 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.[28]

For evaluating model performance, we computed the number of true positives, false positives, false negatives, true negatives, sensitivity (true positive rate), precision (positive predictive value). 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.[29,30] 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 (AUC) and the area under the precision recall curve (AUPRC). To evaluate the statistical difference between the classifiers, we applied a Bayesian metric comparison P(sensitivity),[31] 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(sensitivity) is ≈ 50%.

Results

Median age in the 3913 patients was 70 years (IQR 62-76), 59% were female and 58% had total hip replacement (table 1). Details on prescribed drug types are shown in Appendix 1. Median length of stay 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 positive prediction fraction 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 the full machine-learning (figure 1a) and full logistic regression model leading to this risk-patient selection, 106 and 98 had outcome A, respectively. Correspondingly, the sensitivity and precision were 58.2% and 13.6% for the full machine-learning and 53.8% and 12.5% for the full logistic regression model, respectively. The full machine-learning model 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 positive prediction fractions of 25% and 30%, but with the sensitivity for the full machine-learning model increasing to 64.4% and 69.2% and precision decreasing to 12.0% and 10.7%, respectively (Appendix table 2).

Both the machine-learning model excluding age and age-only model had significantly lower sensitivity than the full machine-learning model (figure 1b). Despite age being the single most important variable (figure 2), the machine-learning model excluding age performed as well as the age-only model.

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 (outcome B) were similar as for outcome A. In general, we found that the full machine-learning model was superior to the others, although the difference were smaller than for outcome A. (Supplemental Digital Content table S1 listing outcome parameters and Supplemental Digital Content 2 figure S1a-b showing distributions and ROC curves for outcome B). 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 Digital Content figure S2a-b showing SHAP values for outcome B). 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 Digital Content 4 figure S3a-d showing SHAP-values of prescriptions of specific drugs for outcome B).

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 outcome A and B, it was less suited for prediction of postoperative medical complications after fast-track total hip and knee replacement 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.[32] 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.[32] 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,[33] 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 UKA[12] and spine surgery,[9] they are not standard in our fast-track protocols and not easy to interpret from a pathophysiological point of view. 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 length of stays or postoperative hemoglobin have been included in other studies[33], we decided against the use of peri- and postoperative data. The same approach has been used by Ramkumar *et al.* who used U.S. National Inpatient Sample data including 15 preoperative variables, to predict length of stay, patient charges and disposition after both TKA[34] and THA.[18] 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 the reason for increased length of stay or non-home discharge.[34] Wei *et al.* used an artificial neural network model to predict same-day discharge after TKA, based on the NSQUIP database from 2018 and found that six of the ten most important variables were the same compared with logistic regression, similar to our findings.[35] However, patients with one-day length of stay were intentionally excluded due to variations in in-patient vs. out-patient registration.[35]

Age has traditionally been a major factor when predicting surgical outcomes 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 the machine-learning model without age being comparable to the age-only model. Note that, although elderly patients had increased risk of postoperative complications, likely related to decline of physical reserves,[36] the use of chronological age alone as a selection criteria for being a “risk-patient” was inferior compared to both machine-learning and logistic regression models incorporating comorbidity and functional status.

We used the SHAP values for estimation of feature importance, thus providing a better understanding of the otherwise “black-box” machine-learning model. The SHAP values showed which variables contribute most to the risk-score predictions.

Our inclusion of specific data on reimbursed prescriptions 6 months prior to surgery based upon the unique Danish registries, unsurprisingly found increased risk-scores with increased number of prescriptions and with the majority being in elderly patients. Similarly, a Canadian study in elective non-cardiac surgery found decreased survival and increased length of stay and readmissions and costs in patients >65 years with polypharmacy.[28] However, this is a complex relationship where some patients benefit from their treatments, while other may suffer from undesirable side-effects. Consequently, the authors cautioned against altering perioperative practices based on current evidence.[28] However, the information from the included prescriptions with SHAP analysis may provide inspiration for new hypothesis-generating studies such as investigation of the potential differences in risk-profile between having preoperative prescribed

VKA and DOAKs. Also, the age-related differences in risk from SSRI's seen in our study could guide further studies on "deprescription".

Another requirement for machine-learning-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 machine-learning algorithm with only the ten most important variables was slightly worse at predicting outcome B than the full logistic regression model. A reason for this could be that when including a length of stay >4 days but without described medical complications, the combination of all variables provide information not available by merely including the ten most important ones. This highlights the need for as much detailed, and preferably non-binary, data as possible to fulfill the true potential of machine-learning algorithms.

Our study has some limitations. First, one of the strengths of machine learning compared to logistic regression is the analysis of multilevel continuous data, whereas we included only a limited number of, often binary, preoperative variables. This could have limited the full realization of our machine learning model. As previously mentioned, we excluded intraoperative information, including type of anesthesia, surgical approach 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.[21] Furthermore, it is doubtful whether the patients used all types of drugs at the time of surgery (e.g. 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 our approach to be superior to depending only on diagnostic codes which often are inaccurate[37] and provide limited details on whether the complication may be attributed to a medical or surgical adverse event. The strengths of our study include the use of national registries with high degree of completion (>99% of all somatic admissions in case of the DNDRP),[38] 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 suggest that machine-learning-algorithms likely 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 benefit from intensified perioperative care.

References

1. Petersen PB, Kehlet H, Jorgensen CC, Lundbeck Foundation Centre for Fast-track H, Knee Replacement Collaborative G: 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. 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
4. Jorgensen CC, Gromov K, Petersen PB, Kehlet H: Influence of day of surgery and prediction of LOS > 2 days after fast-track hip and knee replacement. Acta Orthop 2021; 92: 170-175
5. 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
6. Johns WL, Layon D, Golladay GJ, Kates SL, Scott M, Patel NK: Preoperative Risk Factor Screening Protocols in Total Joint Arthroplasty: A Systematic Review. J Arthroplasty 2020; 35: 3353-3363
7. 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
8. Shah A, Memon M, Kay J, Wood TJ, Tushinski DM, Khanna V, McMaster Arthroplasty Collective g: Preoperative Patient Factors Affecting Length of Stay following Total Knee Arthroplasty: A Systematic Review and Meta-Analysis. J Arthroplasty 2019; 34: 2124-2165 e1
9. 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
10. 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
11. 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 2022; 35: 7-14
12. 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
13. 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
14. 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
15. 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

16. 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
17. 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
18. 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
19. 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
20. Johannesdottir KB, Kehlet H, Petersen PB, Aasvang EK, Sørensen HBD, Jørgensen CC: 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 2022; 93: 117-123
21. 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
22. 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
23. Olczak J, Pavlopoulos J, Prijs J, Ijpma FFA, Doornberg JN, Lundstrom 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; 92: 513-525
24. 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
25. 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
26. 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
27. Lundberg SMLSI: A Unified Approach to Interpreting Model Predictions. Edited by Guyon I. Adv Neural Inf Process Syst [Internet], Curran Associates, Inc., 2017
28. 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
29. Chicco D: Ten quick tips for machine learning in computational biology. BioData Mining 2017; 10: 35 (2017)
30. Chicco D, Totsch 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 Mining 2021; 14: 13 (2021)
31. Totsch N, Hoffmann D: Classifier uncertainty: evidence, potential impact, and probabilistic treatment. PeerJ Comput Sci 2021; 7: e398

32. 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
33. 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
34. 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
35. 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
36. 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. Anaesthesia 2014; 69 Suppl 1: 81-98
37. Bedard NA, Pugely AJ, McHugh MA, Lux NR, Bozic KJ, Callaghan JJ: Big Data and Total Hip Arthroplasty: How Do Large Databases Compare? J Arthroplasty 2018; 33: 41-45.e3
38. 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

Figure legends

Figure 1a-b

1a) Distribution of full machine learning model risk scores for patients +/- outcome A. 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), machine learning excluding age (MLM -age) 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 outcome A (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 outcome A, where positive increase and negative values decrease the risk score. The color is related to the value of the variable with blue being lowest and red highest and each dot represents a patient.

Figure 3a-d

SHAP scatter-plot on the contributions to the full machine-learning model on outcome A (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 Ca^{2+}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.

Table 1. patient demographics with and without outcome A (length of stay >4 days or readmissions due to "medical" morbidity) 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 prescriptions[1] (SD)	3.0 (1.0-4.0)	2.0 (0.0-3.0)
female gender	755 (64.0)	12133 (58.2)
Total hip replacement	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	552 (46.8)	4398 (21.5)
missing	29 (2.5)	359 (1.7)
living alone	578 (49.0)	6717 (32.2)
with others	571 (48.4)	13869 (66.6)
institution	24 (2.0)	113 (0.5)
missing	7 (0.6)	138 (0.7)
hemoglobin	8.2 (7.7-8.8)	8.6 (8.1-9.2)
missing	11 (0.9)	314 (1.5)
>2 units of alcohol/day	79 (6.7)	1589 (7.6)
missing	10 (0.8)	174 (0.8)
active smoker	130 (11.0)	2751 (13.2)
missing	11 (0.9)	141 (0.7)
cardiac disease	306 (25.9)	2750 (13.2)
missing	8 (0.8)	153 (0.7)
hypercholesterolemia	467 (39.6)	6062 (29.1)
missing	8 (0.7)	120 (0.6)
hypertension	738 (62.5)	10141 (48.7)
missing	64 (5.4)	663 (3.2)
pulmonary disease	182 (15.4)	1841 (8.8)
missing	5 (0.4)	96 (0.5)
previous cerebral attack	165 (14.0)	1086 (5.2)
missing	25 (2.1)	282 (1.4)
previous VTE	133 (11.3)	1481 (7.1)
missing	26 (2.2)	325 (1.6)
malignancy (undefined)	557 (47.2)	8843 (42.4)
previous radically treated malignancy	127 (10.8)	2065 (9.9)
missing	14 (1.2)	162 (0.8)
chronic kidney disease	50 (4.2)	273 (1.3)
missing	35 (3.0)	292 (1.4)
family member with VTE	155 (13.1)	2510 (12.0)
missing	1190 (16.1)	2569 (12.3)
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	468 (39.7)	9340 (44.8)

uncertain about being rested	48 (4.1)	809 (3.9)
missing	105 (8.9)	1230 (5.9)
psychiatric disorder	156 (13.2)	1590 (7.6)
missing	62 (5.3)	703 (3.4)
Characteristic based on combination of questionnaire and DNDRP		
diabetes		
diet treated diabetes[2]	29 (2.5)	274 (1.3)
oral antidiabetics	137 (11.6)	1448 (6.9)
insulin treated diabetes[3]	60 (5.1)	413 (2.0)
missing	7 (0.6)	98 (0.5)

SD: standard deviation VTE: venous thromboembolic event DNDRP: Danish National Database of Reimbursed Prescriptions.

[1]Antirheumatica, steroids, anticoagulants, cardiac, cholesterol lowering, respiratory and psychotropic drugs.

[2]Reported diabetes but no registered prescriptions [3] +/- oral antidiabetics

table2

Table 2: Performance of the six different models with a predefined positive prediction fraction of 20% for outcome A

Positive prediction fraction 20%	TP	FP	FN	TN	sensitivity	precision	MCC	AUROC	AUPRC	P (sensitivity)
Full machine-learning model	106	676	76	3055	58.2%	13.6%	21.1%	76.3%	15.5%	-
Full logistic regression model	98	684	84	3047	53.8%	12.5%	18.7%	74.5%	15.7%	19.7%
Parsimonious machine-learning model	100	682	82	3049	54.9%	12.8%	19.3%	75.9%	17.3%	26.1%
Parsimonious logistic regression model	95	687	87	3045	52.2%	12.1%	17.8%	73.7%	13.6%	12.4%
machine-learning model excluding age	88	694	94	3037	48.4%	11.3%	15.7%	72.3%	13.6%	3.1%
Age-only model	87	676	95	3055	47.8%	11.4%	15.8%	69.7%	12.1%	2.3%

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

Appendix table 1

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

	+Outcome A	-Outcome A
Anticoagulants		
none	679 (57.5)	15844 (76.0)
VKA	106 (9.0)	750 (3.6)
Heparin & Acetylsalicylic acid	0 (0.0)	7 (0.0)
DOAC	48 (4.1)	659 (3.2)
Acetylsalicylic acid	205 (17.4)	2492 (12.0)
Dipyradimol	5 (0.4)	29 (0.1)
ADP-antagonist	75 (6.4)	569 (2.7)
Acetylsalicylic acid & Dipyradimol	17 (1.4)	168 (0.8)
VKA & Acetylsalicylic acid	10 (0.8)	78 (0.4)
DOAC & Acetylsalicylic acid	6 (0.5)	41 (0.2)
VKA & ADP-antagonist	4 (0.3)	11 (0.1)
DOAC & ADP-antagonist	3 (0.3)	14 (0.1)
VKA & Heparin	1 (0.1)	21 (0.1)
DOAC & Acetylsalicylic acid & ADP-antagonist	1 (0.1)	3 (0.0)
Acetylsalicylic acid & ADP-antagonist	18 (1.5)	132 (0.6)
Acetylsalicylic acid & ADP-antagonist & Heparin	1 (0.1)	12 (0.1)
Acetylsalicylic acid & ADP-antagonist & Dipyradimol	1 (0.1)	7 (0.0)
Cardiac prescriptions		
none	321 (27.2)	9200 (44.2)
diuretics	77 (6.5)	1184 (5.7)
angiotensin-II/ACE-inhibitors	132 (11.2)	2683 (12.9)
Ca^{2+} antagonists	55 (4.7)	773 (3.7)
β-blocker	29 (2.5)	559 (2.7)
nitrates	1 (0.1)	18 (0.1)
other antihypertensives	0 (0.0)	12 (0.1)
other types of medication for IHD	2 (0.2)	21 (0.1)
2 antihypertensives	177 (15.0)	2696 (12.9)
β-blocker & 1 antihypertensive[1]	92 (8.1)	1069 (5.1)
3 antihypertensives	50 (4.2)	548 (2.6)
β-blocker & 2 antihypertensives[1]	95 (8.1)	975 (4.7)
β-blocker & 3 antihypertensives[1]	25 (2.1)	265 (1.3)
4 antihypertensives	2 (0.2)	18 (0.1)
β-blocker & 4 antihypertensives	2 (0.2)	19 (0.1)
other antihypertensive & antihypertensives[1]	9 (0.8)	87 (0.4)
nitrates & any hypertensive	49 (4.2)	331 (1.6)
other drugs for IHD & any antihypertensive and/or nitrate	5 (0.4)	15 (0.1)
other antiarrhythmics & any antihypertensives	57 (4.8)	364 (1.7)
Anticholesterols		
none	708 (60.0)	14719 (70.6)
statins	457 (38.7)	5866 (28.2)
other anti-lipids	7 (0.6)	135 (0.6)
Statins +other anti-lipids	8 (0.7)	117 (0.6)

Systemic steroids	123 (10.4)	1149 (5.5)
Antirheumatics		
none	1143 (96.9)	20388 (97.8)
disease-modifying antirheumatic drugs	37 (3.1)	446 (2.1)
other antirheumatics	0 (0.0)	3 (0.0)
Respiratory prescriptions		
none	1000 (84.7)	18754 (90.0)
SABA	13 (1.1)	276 (1.3)
LABA or LAMA	19 (1.6)	217 (1.0)
inhalation steroid only	8 (0.7)	211 (1.0)
SABA & Ipratropium (+/- others)	6 (0.5)	18 (0.1)
LABA & steroid	45 (3.8)	474 (2.3)
LABA & LAMA & steroid	19 (1.6)	122 (0.6)
LAMA & steroid	0 (0.0)	11 (0.1)
LABA & LAMA	7 (0.6)	80 (0.4)
other pulmonary drugs	3 (0.3)	32 (0.2)
other pulmonary drugs & steroid	9 (0.8)	98 (0.5)
SABA & LABA or LAMA	6 (0.5)	96 (0.5)
SABA & LABA or LAMA & steroid	45 (3.8)	448 (2.2)
Psychotropic prescriptions		
none	952 (80.7)	18657 (89.5)
SSRI/SNRI/NaRI	100 (8.5)	1164 (5.6)
other antidepressants	1 (0.1)	17 (0.1)
antipsychotics	8 (0.7)	116 (0.6)
benzodiazepines[2]	0 (0.0)	7 (0.0)
anti-cholinergics or memantine	6 (0.5)	27 (0.1)
anti-ADHD drugs	1 (0.1)	10 (0.0)
NaSSA	25 (2.1)	184 (0.9)
other psychotropics	28 (2.4)	182 (0.9)
SSRI + other antidepressants	4 (0.3)	6 (0.0)
SSRI + NaSSA	8 (0.7)	94 (0.5)
SRRI + antipsychotics	11 (0.9)	87 (0.4)
SRRI + other psychotropics	7 (0.6)	84 (0.4)
benzodiazepines + any psychotropic	3 (0.3)	12 (0.1)
antipsychotics + any psychotropic	20 (1.7)	149 (0.7)
anti-ADHD + any psychotropic	0 (0.0)	14 (0.1)
NaSSA + any psychotropic	4 (0.3)	18 (0.1)
other psychotropics + any specified psychotropic	2 (0.2)	9 (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

[1]either diuretics, ACE/ANG-II inhibitors or Ca^{2+}antagonists [2] likely underreported due to limited general reimbursement for benzodiazepines in Denmark

Appendix table 2

Performance of the six different models with a predefined positive prediction fraction of 25% and 30% for outcome A (LOS >4 days or readmission due to “medical” morbidity.

Positive prediction fraction 25%	TP	FP	FN	TN	sensitivity	precision	MCC	AUC	AUPRC	P (sensitivity)
Full machine-learning model	117	861	65	2870	64.3%	12.0%	20.0%	76.3%	15.5%	-
Full logistic regression model	110	868	72	2863	60.4%	11.2%	18.1%	74.5%	15.7%	23.1%
Parsimonious machine-learning model	115	863	67	2868	63.2%	11.8%	19.5%	75.9%	17.3%	41.2%
Parsimonious logistic regression model	106	872	76	2859	58.2%	10.8%	17.0%	73.4%	15.5%	11.8%
machine-learning model excluding age	106	872	76	2859	58.2%	10.8%	17.0%	72.3%	13.6%	11.8%
Age-model	94	824	88	2907	51.6%	10.2%	14.7%	69.7%	12.2%	0.7%
Positive prediction fraction 30%	**TP**	**FP**	**FN**	**TN**	**sensitivity**	**precision**	**MCC**	**AUC**	**AUPRC**	**P (sensitivity)**
Full machine-learning model	126	1047	56	2684	69.2%	10.7%	18.9%	76.3%	15.5%	-
Full logistic regression model	120	1053	62	2678	65.9%	10.2%	17.3%	74.5%	15.7%	25.2%
Parsimonious machine-learning model	124	1049	58	2682	68.1%	10.6%	18.4%	75.9%	17.3%	40.8%
Parsimonious logistic regression model	115	1058	67	2673	63.2%	9.8%	16.0%	73.7%	15.5%	11.1%
machine-learning model excluding age	116	1057	66	2674	63.7%	9.9%	16.3%	72.3%	13.6%	13.8%
Age-model	100	955	82	2776	54.9%	9.5%	13.9%	69.7%	12.2%	0.2%

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

Fig1

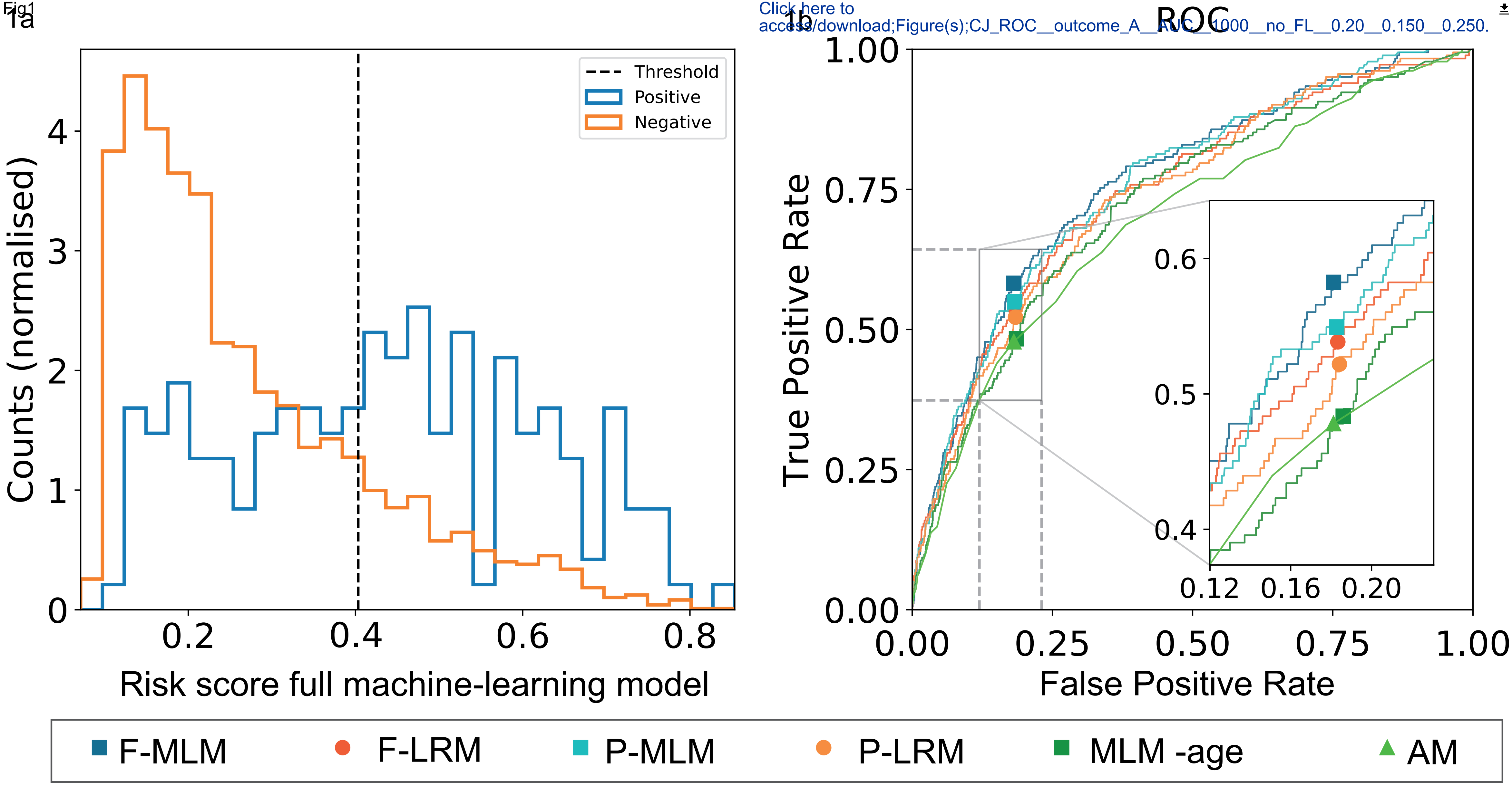
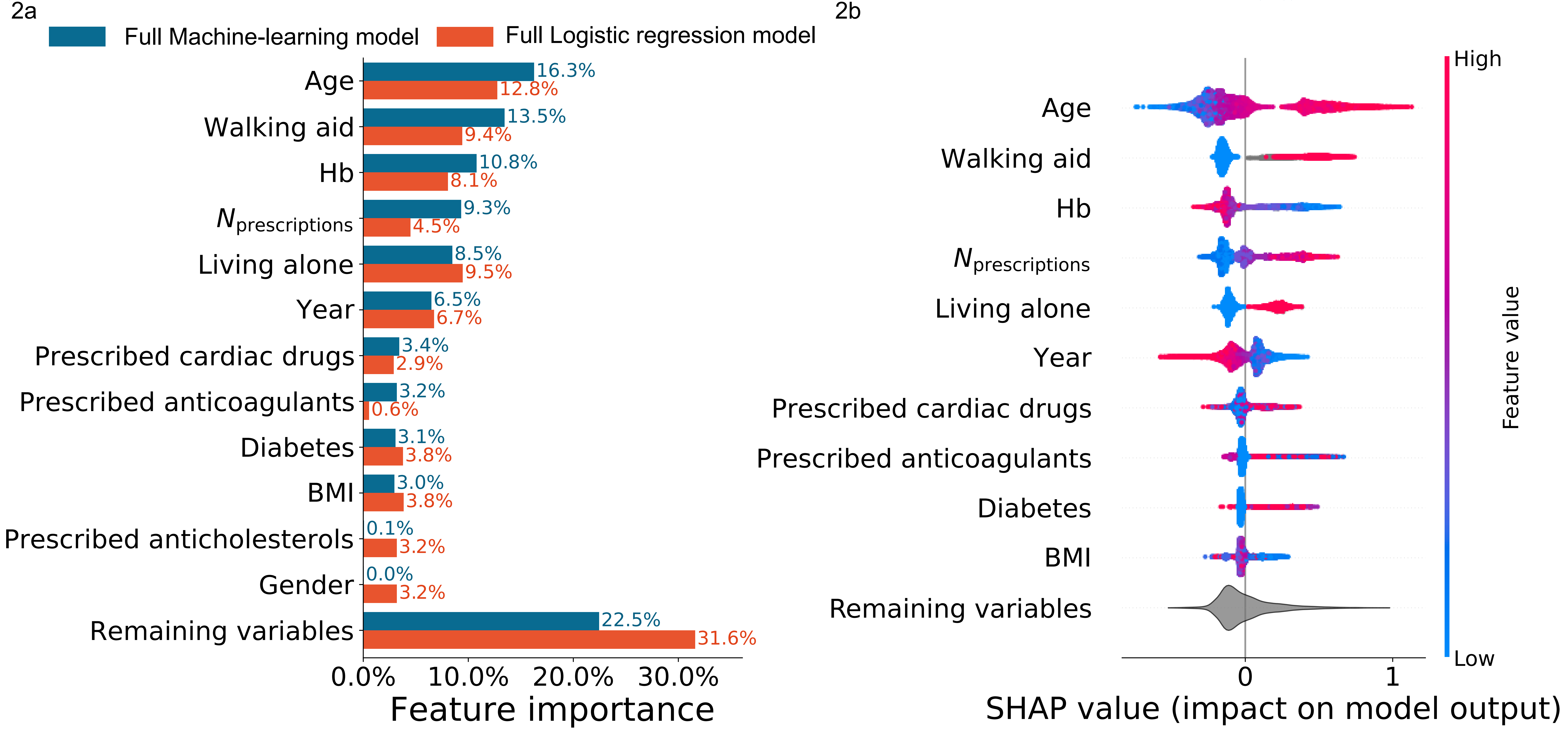
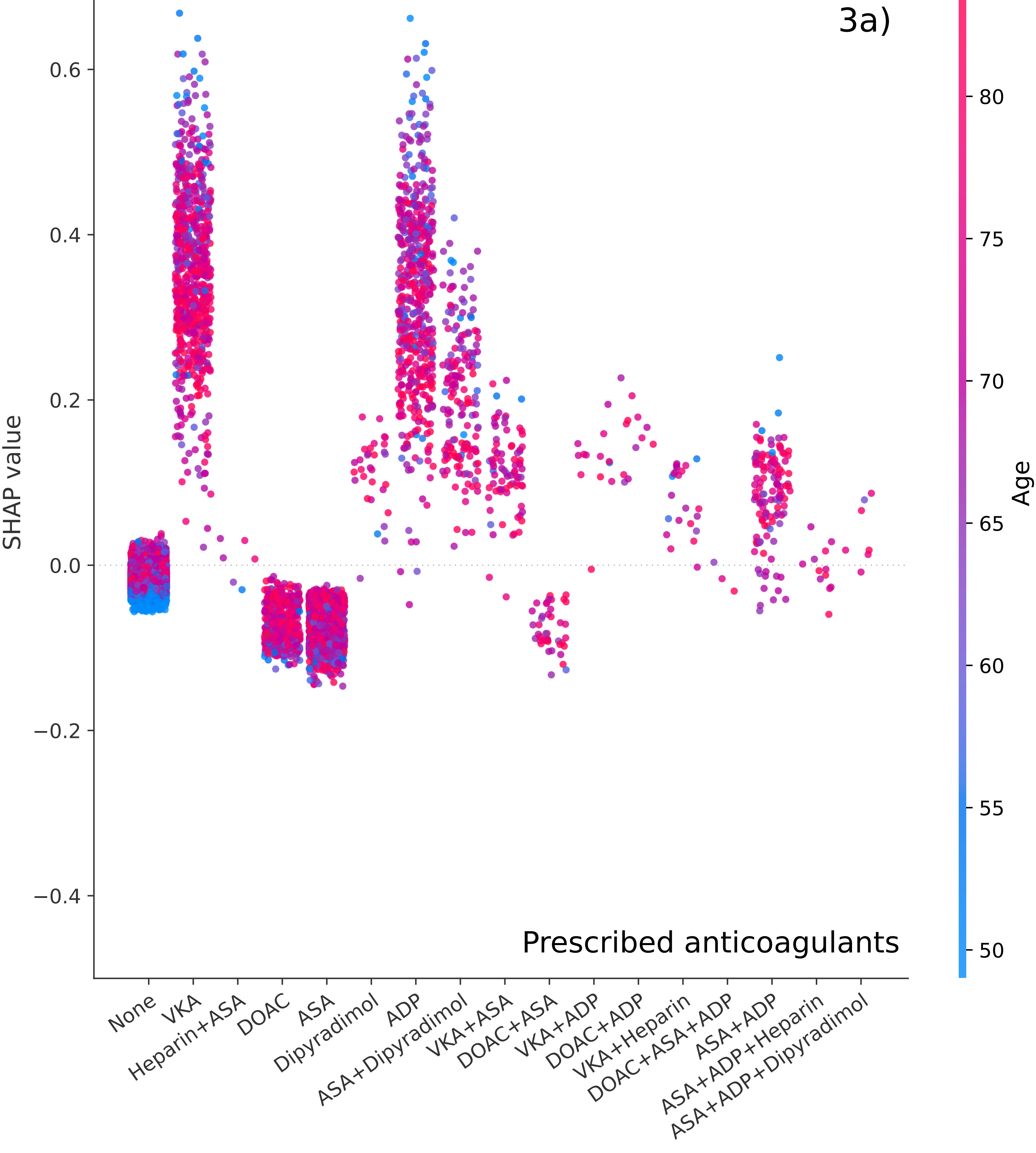

1a
Threshold
Positive
Negative
Counts (normalised)
0
1
2
3
4
0.2
0.4
0.6
0.8
Risk score full machine-learning model
1b
ROC
True Positive Rate
False Positive Rate
0.00
0.25
0.50
0.75
1.00
0.4
0.5
0.6
0.12
0.16
0.20
F-MLM
F-LRM
P-MLM
P-LRM
MLM -age
AM

Fig2



2a
Full Machine-learning model
Full Logistic regression model
Age
16.3%
12.8%
Walking aid
13.5%
9.4%
Hb
10.8%
8.1%
$N_{\text{prescriptions}}$
9.3%
4.5%
Living alone
8.5%
9.5%
Year
6.5%
6.7%
Prescribed cardiac drugs
3.4%
2.9%
Prescribed anticoagulants
3.2%
0.6%
Diabetes
3.1%
3.8%
BMI
3.0%
3.8%
Prescribed anticholesterols
0.1%
3.2%
Gender
0.0%
3.2%
Remaining variables
22.5%
31.6%
0.0%
10.0%
20.0%
30.0%
Feature importance
2b
Age
Walking aid
Hb
$N_{\text{prescriptions}}$
Living alone
Year
Prescribed cardiac drugs
Prescribed anticoagulants
Diabetes
BMI
Remaining variables
0
1
SHAP value (impact on model output)
High
Feature value
Low

fig3a



3a)
SHAP value
0.6
0.4
0.2
0.0
−0.2
−0.4
Prescribed 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
Age
80
75
70
65
60
55
50

fig3b

Click here to access/download;Figure(s);Fig3b.pdf

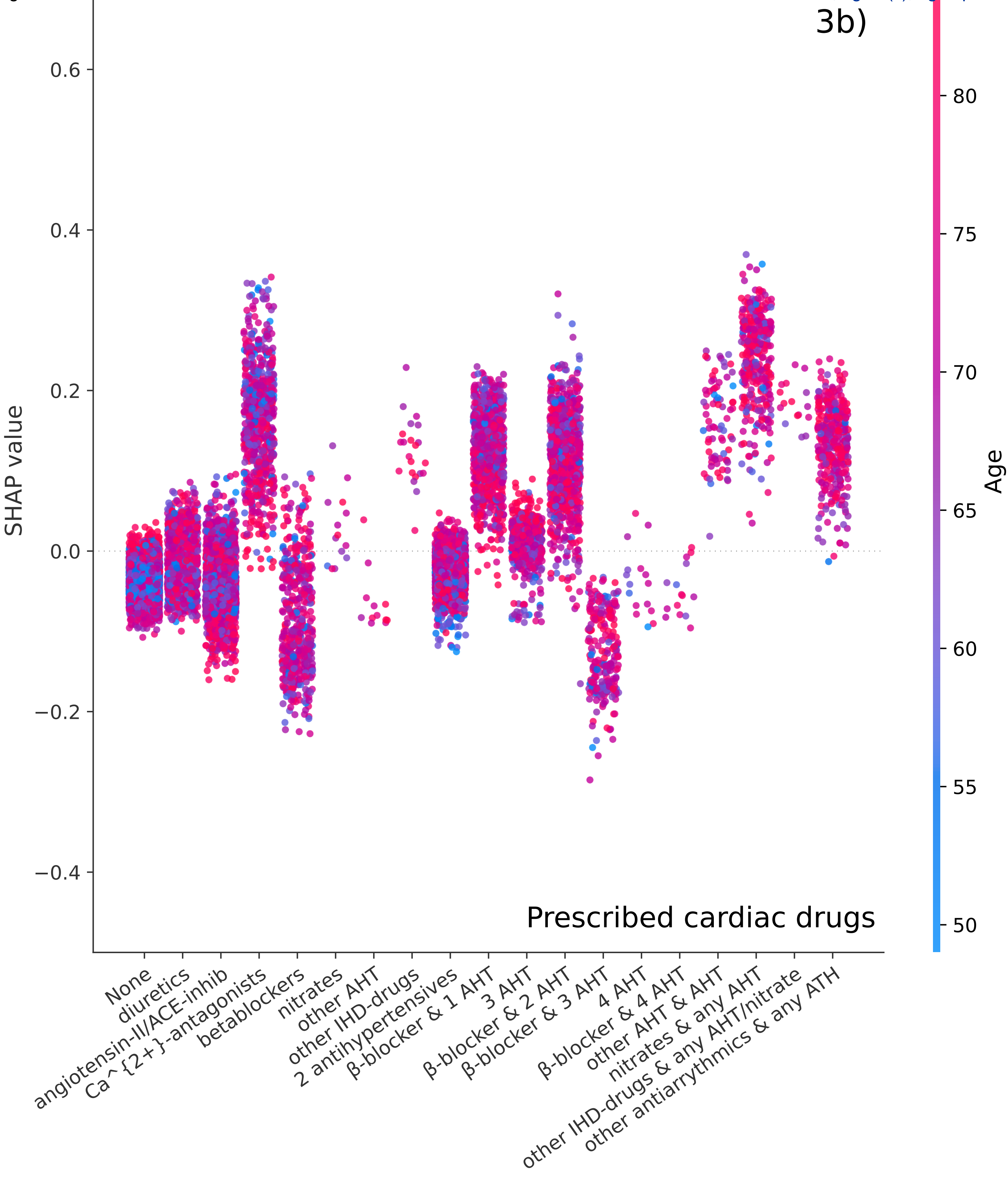


fig3c

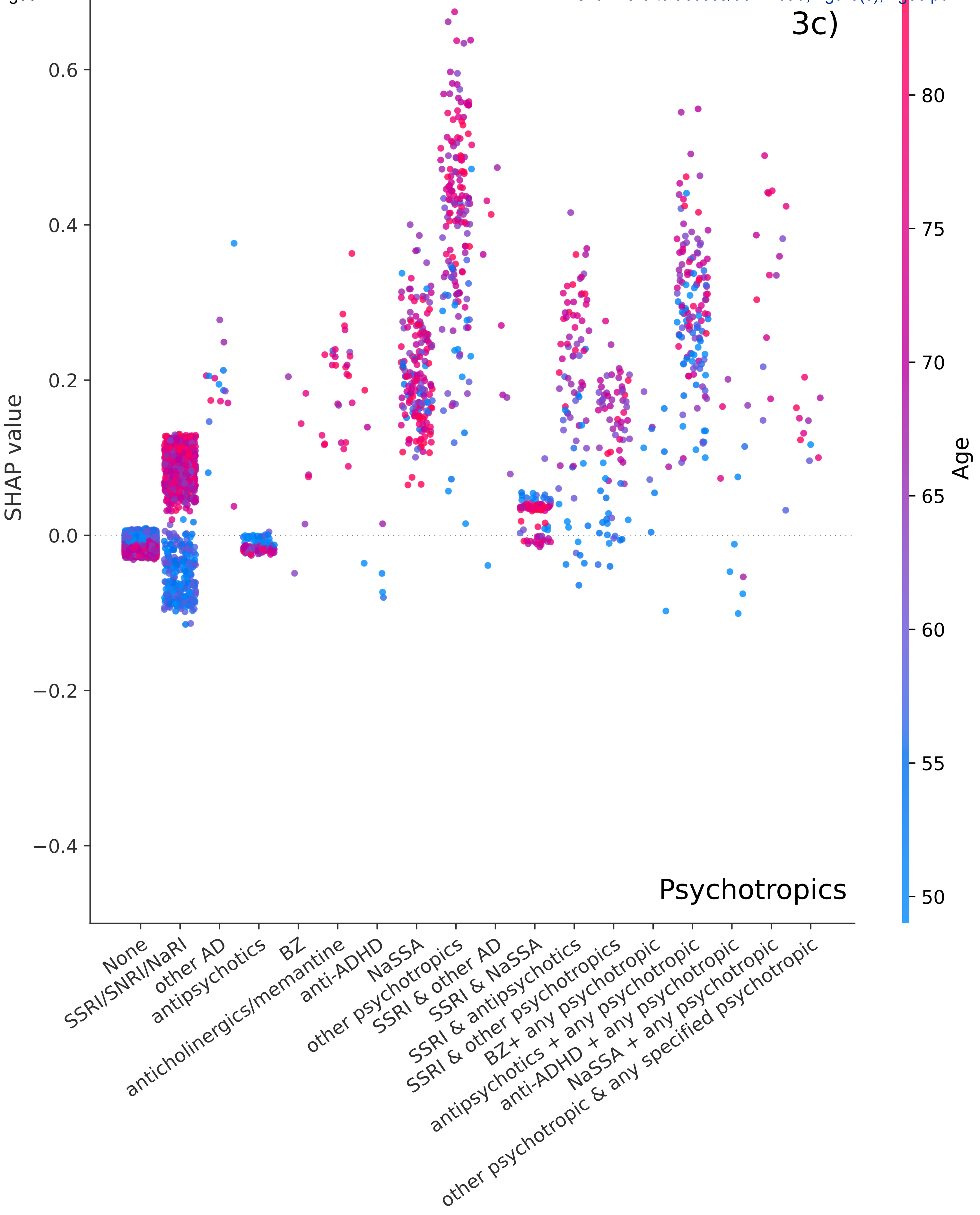
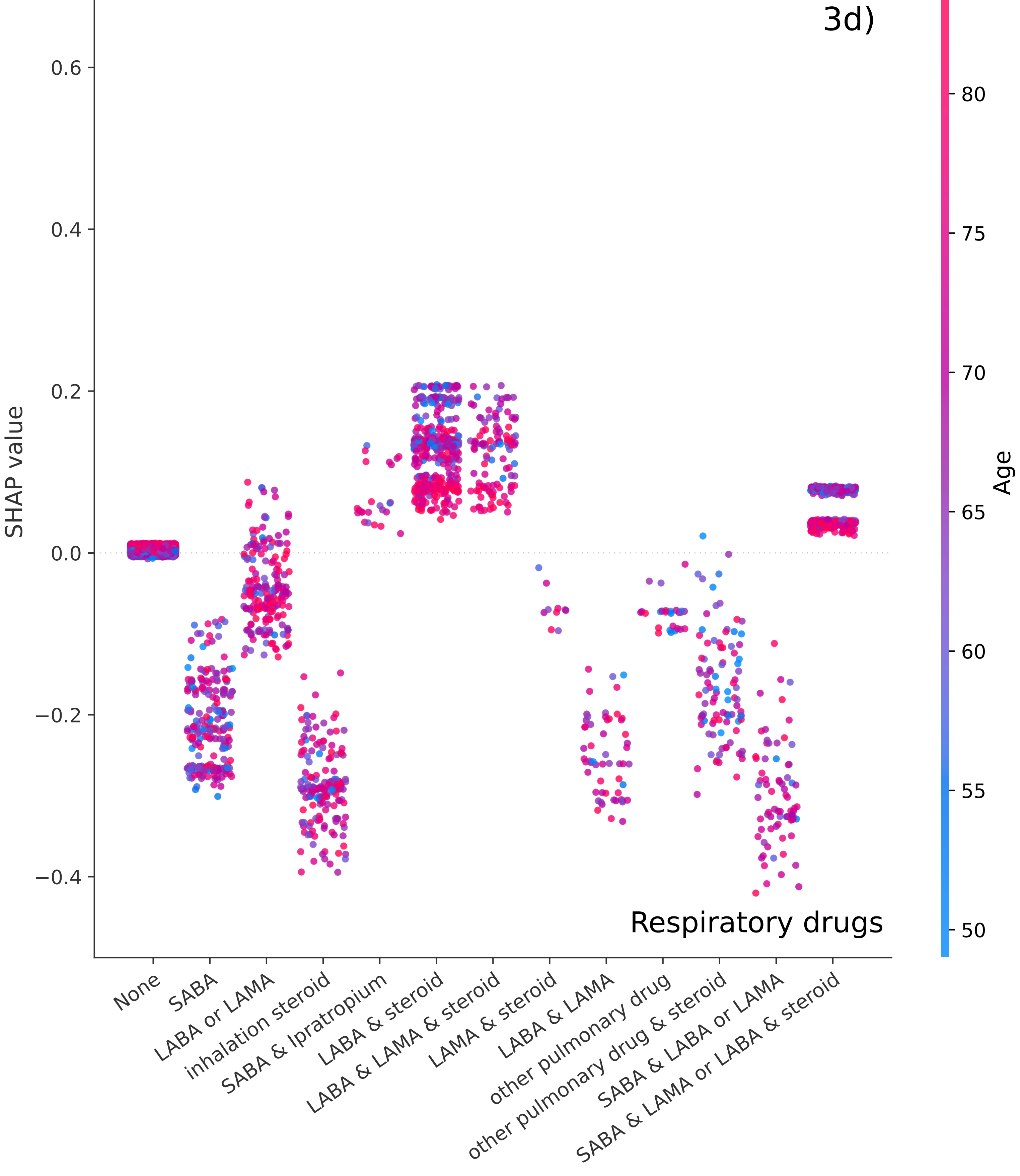

3c)
SHAP value
0.6
0.4
0.2
0.0
−0.2
−0.4
Psychotropics
Age
80
75
70
65
60
55
50
None
SSRI/SNRI/NaRI
other AD
antipsychotics
BZ
anticholinergics/memantine
anti-ADHD
NaSSA
other psychotropics
SSRI & other AD
SSRI & NaSSA
SSRI & antipsychotics
SSRI & other psychotropics
BZ+ any psychotropic
antipsychotics + any psychotropic
anti-ADHD + any psychotropic
NaSSA + any psychotropic
other psychotropic & any specified psychotropic

fig3d


3d)
0.6
0.4
0.2
0.0
−0.2
−0.4
SHAP value
Respiratory drugs
None
SABA
LABA or LAMA
inhalation steroid
SABA & Ipratropium
LABA & steroid
LABA & LAMA & steroid
LAMA & steroid
LABA & LAMA
other pulmonary drug
other pulmonary drug & steroid
SABA & LABA or LAMA
SABA & LAMA or LABA & steroid
80
75
70
65
60
55
50
Age

Supplemental Digital Content 1

Table S1: Performance of different models for Outcome B (Los >4 days or readmissions due to “medical” morbidity or LOS >4 days but without recorded morbidity)

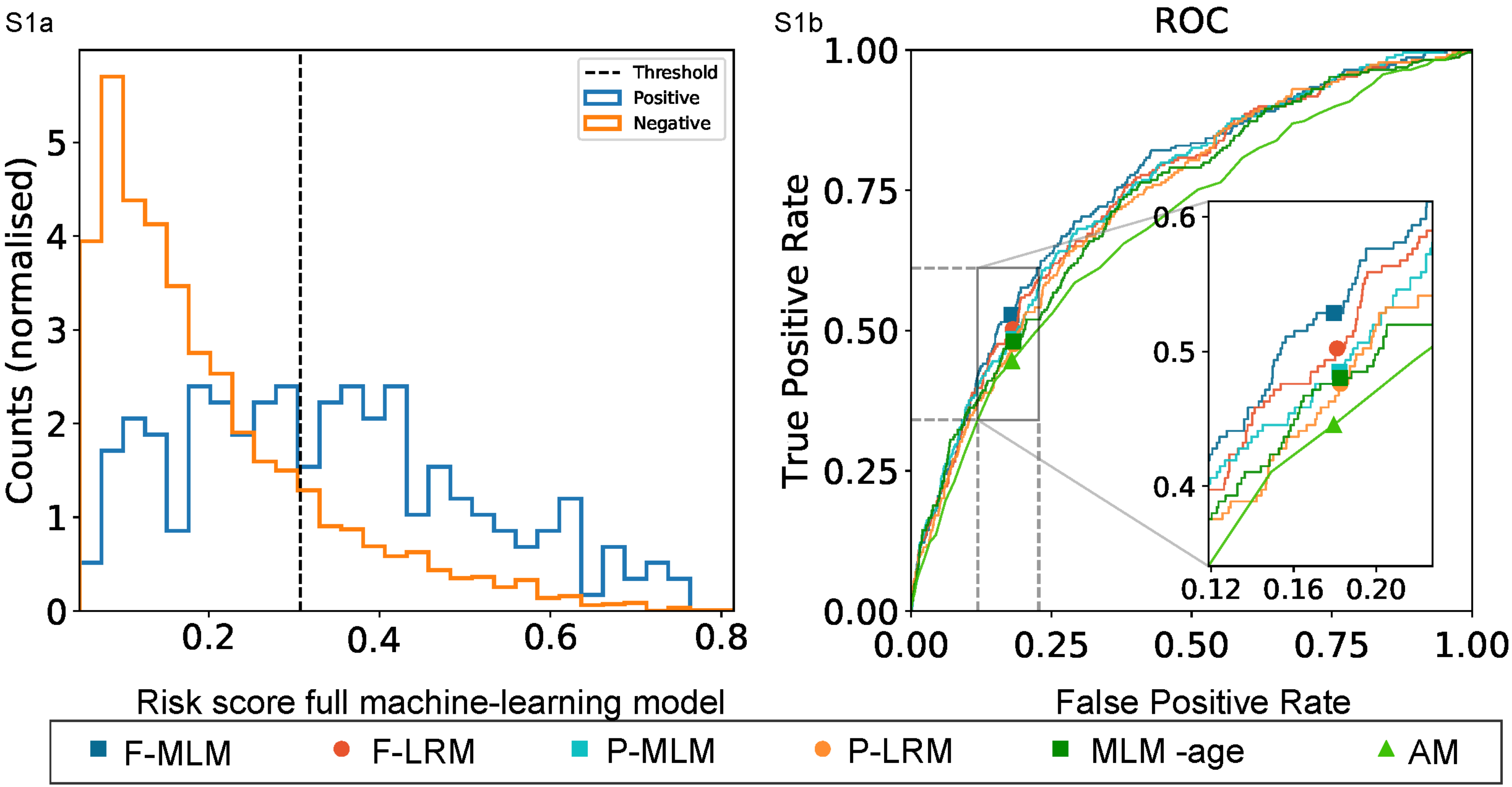
Positive prediction fraction 20%	TP	FP	FN	TN	sensitivity	precision	MCC	AUC	AUPRC	P (sensitivity)
Full machine-learning model	121	661	108	3023	52.8%	15.5%	20.5%	75.3%	17.1%	-
Full logistic regression model	115	667	114	3017	50.2%	14.7%	18.9%	74.1%	16.7%	28.3%
Parsimonious machine-learning model	111	671	118	3013	48.4%	14.2%	17.8%	74.4%	16.8%	17.2%
Parsimonious logistic regression model	109	673	120	3011	47.6%	13.9%	17.2%	73.1%	16.8%	12.9%
machine-learning model excluding age	110	672	119	3012	48.0%	14.1%	17.5%	72.8%	16.9%	15.1%
Age-model	102	661	127	3023	44.5%	13.4%	15.8%	68.7%	13.4%	3.8%
Positive prediction fraction 25%	**TP**	**FP**	**FN**	**TN**	**sensitivity**	**precision**	**MCC**	**AUC**	**AUPRC**	**P (sensitivity)**
Full machine-learning model	140	838	89	2846	61.1%	14.3%	20.8%	75.3%	17.1%	-
Full logistic regression model	136	842	93	2842	59.4%	13.9%	19.8%	74.1%	16.7%	35.3
Parsimonious machine-learning model	134	844	95	2840	58.5%	13.7%	19.3%	74.4%	16.8%	28.3
Parsimonious logistic regression model	125	853	104	2831	54.6%	12.8%	17.0%	73.1%	16.8%	7.8
machine-learning model excluding age	121	857	108	2827	52.8%	12.4%	16.0%	72.8%	16.9%	3.6
Age-model	113	805	116	2879	49.3%	12.3%	15.2%	68.7%	13.4%	0.5
Positive prediction fraction 30%	**TP**	**FP**	**FN**	**TN**	**sensitivity**	**precision**	**MCC**	**AUC**	**AUPRC**	**P (sensitivity)**
Full machine-learning model	153	1020	76	2664	66.8%	13.0%	20.0%	75.3%	17.1%	-
Full logistic regression model	147	1026	82	2658	64.2%	12.5%	18.6%	74.1%	16.7%	27.9

Parsimonious machine-learning model	147	1026	82	2658	64.2%	12.5%	18.6%	74.4%	16.8%	27.7
Parsimonious logistic regression model	145	1028	84	2656	63.3%	12.4%	18.1%	73.1%	16.8%	21.6
machine-learning model excluding age	140	1033	89	2651	61.1%	11.9%	17.0%	72.8%	16.9%	10.2
Age-model	122	933	107	2751	53.3%	11.6%	14.8%	69.8%	13.4%	0.1

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

Supplemental Digital Content 2

Figure S1a-b

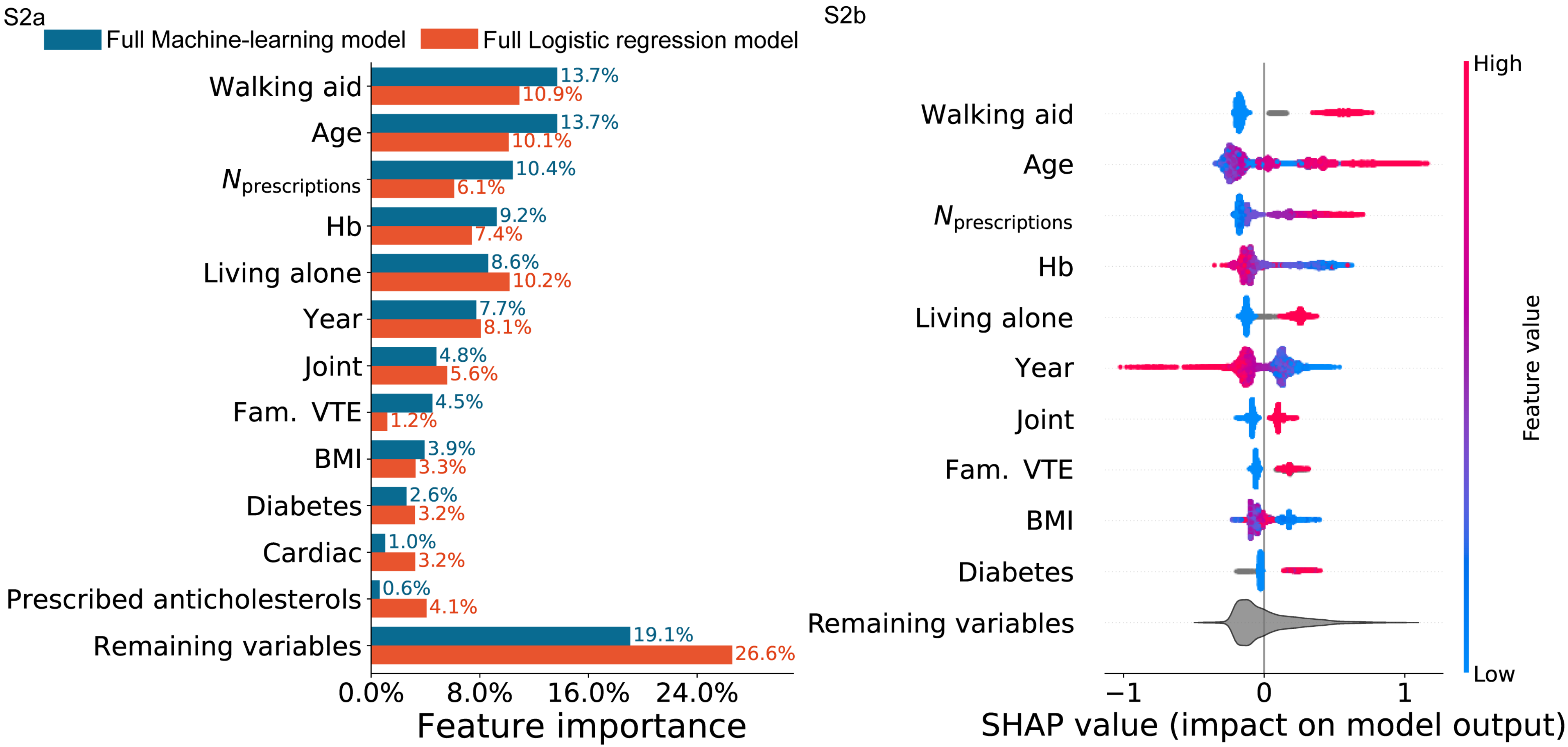


S1a) Distribution of full machine learning model risk scores for patients +/- outcome B(LOS >4 days or readmissions due to “medical” morbidity or LOS >4 days with no recorded morbidity). The dashed line marks the classification threshold of 20% positive prediction fraction.

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), machine learning excluding age (MLM -age) and the age-only model (AM).

Supplemental Digital Content 3

Figure S2a-b

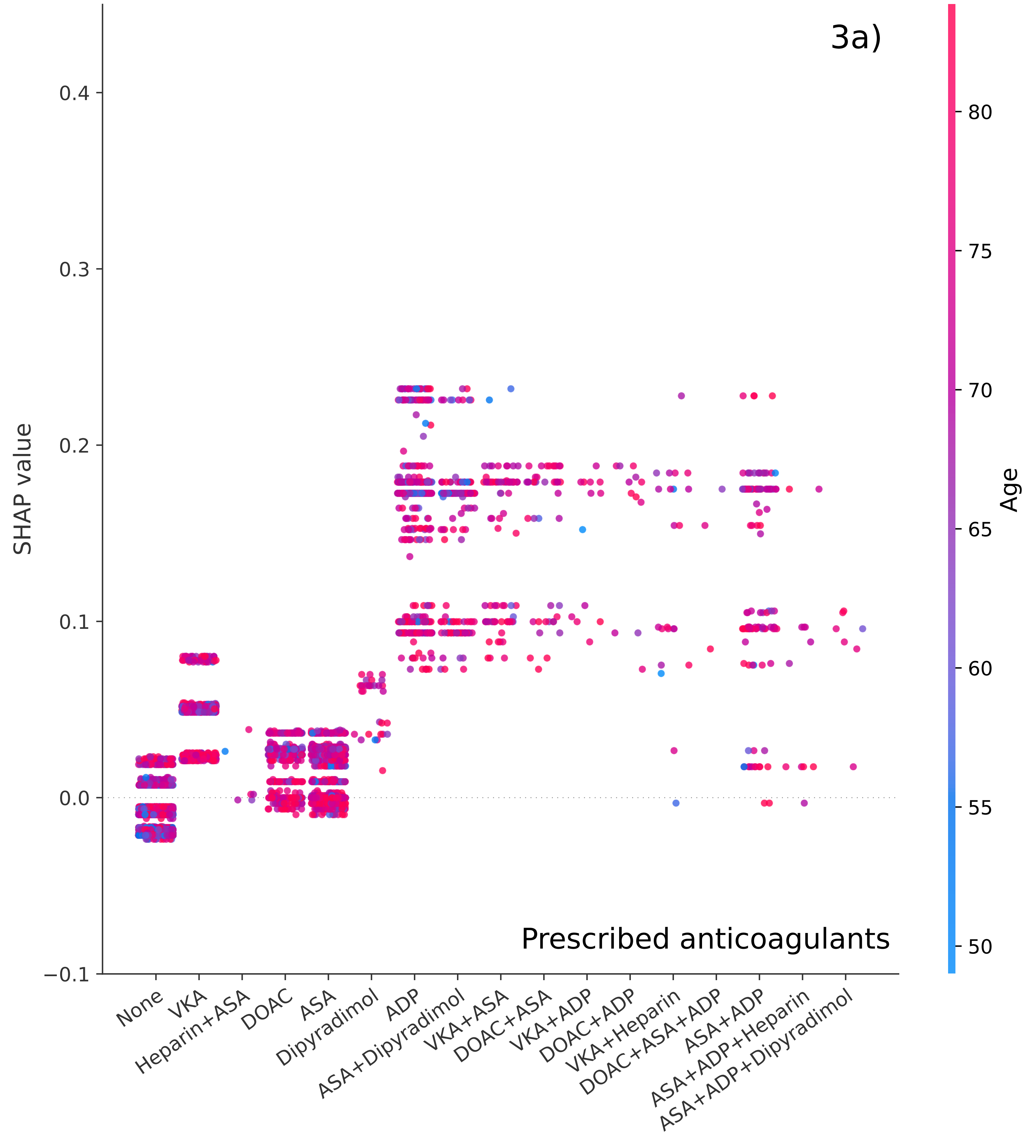


S2a) The overall importance of the 10 most important variables measured by the SHAP-values for the full machine-learning and full logistic regression models for outcome B (LOS >4 days or readmissions due to “medical” morbidity or LOS >4 days with no recorded morbidity).
Only the importance of prescribed 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 where values increase while negative values decrease the risk score. The color is related to the value of the variable with blue being lowest and red highest and each dot represents a patient.

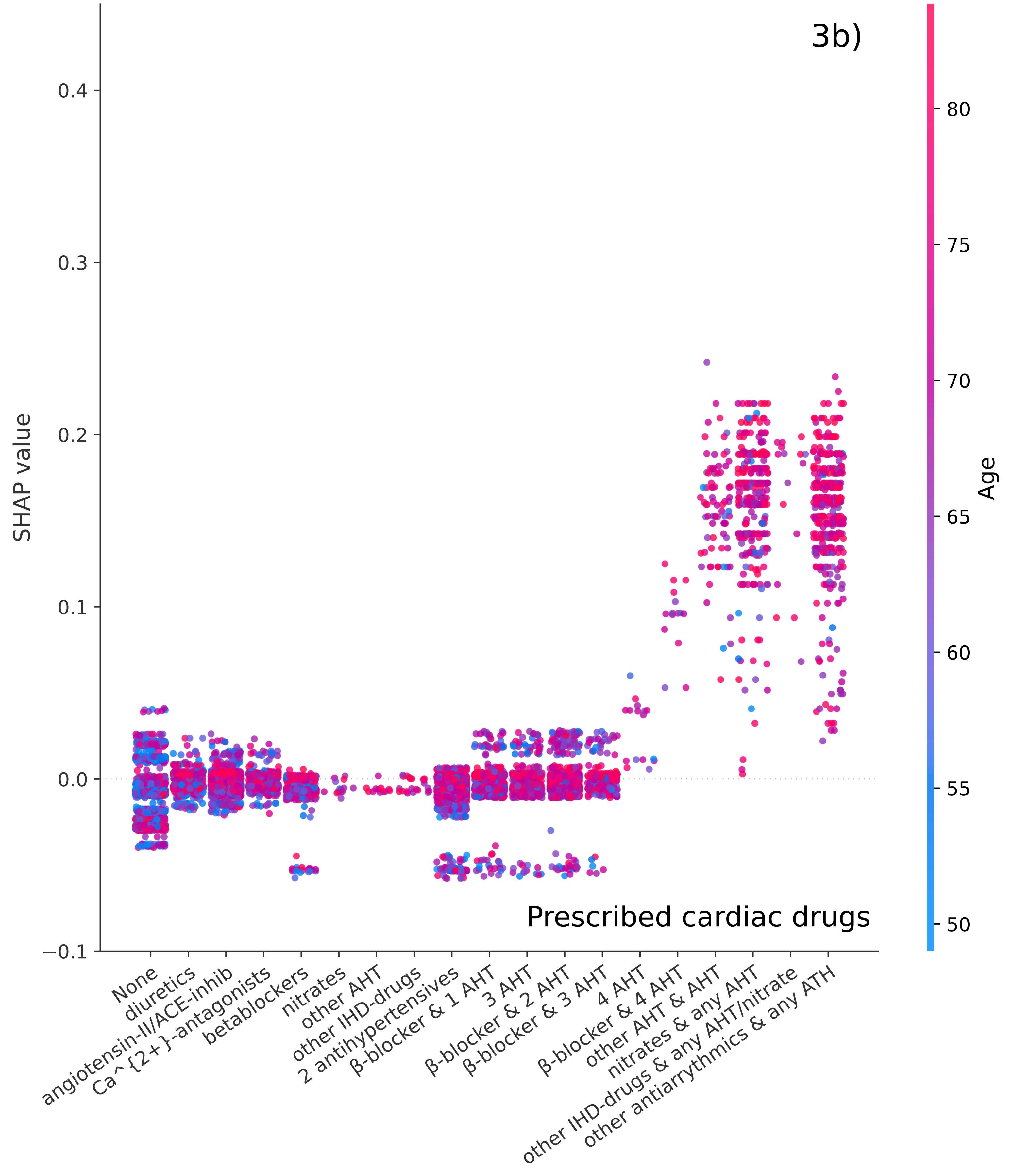
Supplemental Digital Content 4

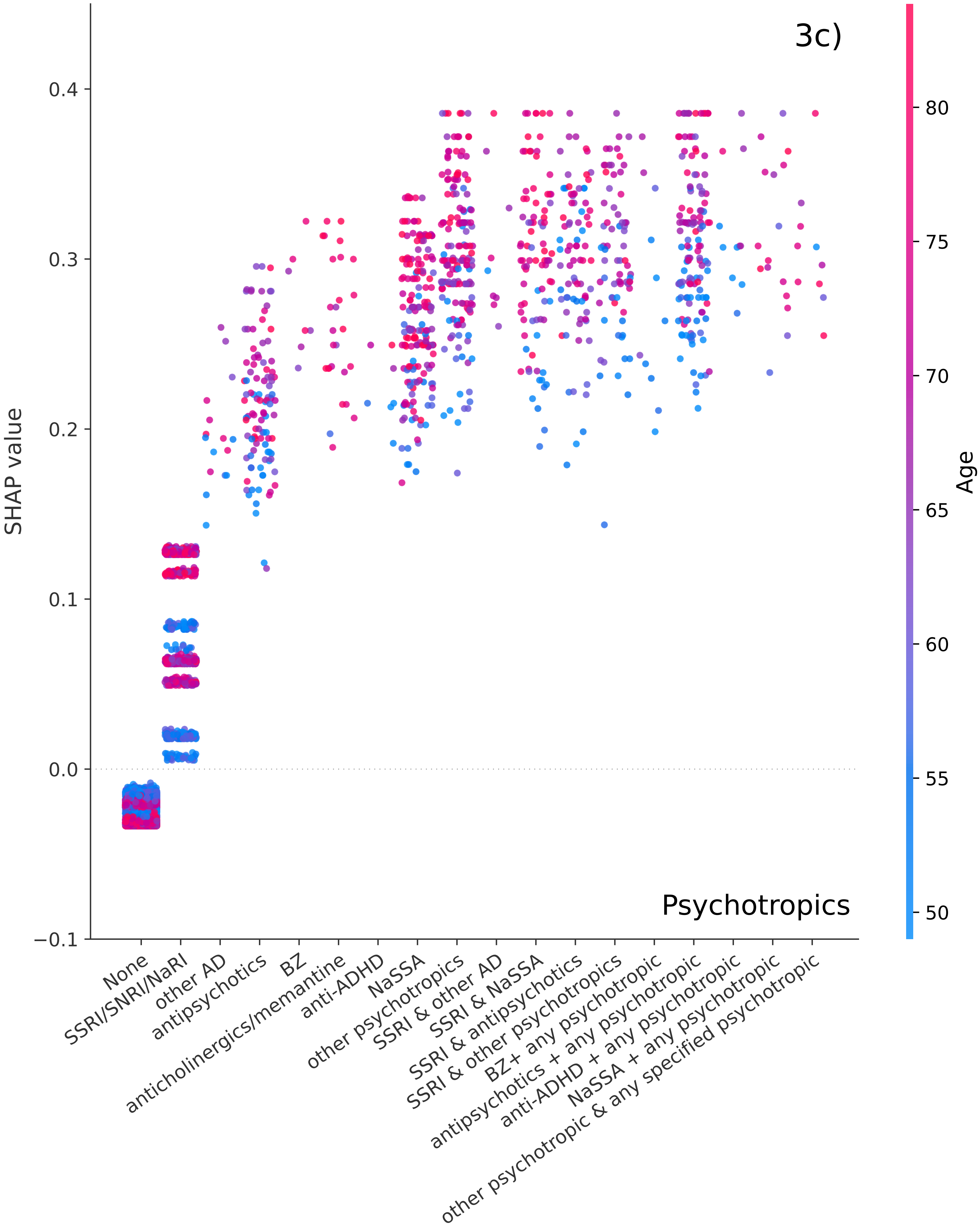
Figure S3a-d
SHAP scatter-plot on the contributions to the full machine-learning model on outcome B (LOS >4 days or readmission due to “medical” morbidity), for individual types of prescribed anticoagulants, cardiac drugs, psychotropics and respiratory drugs stratified by age.

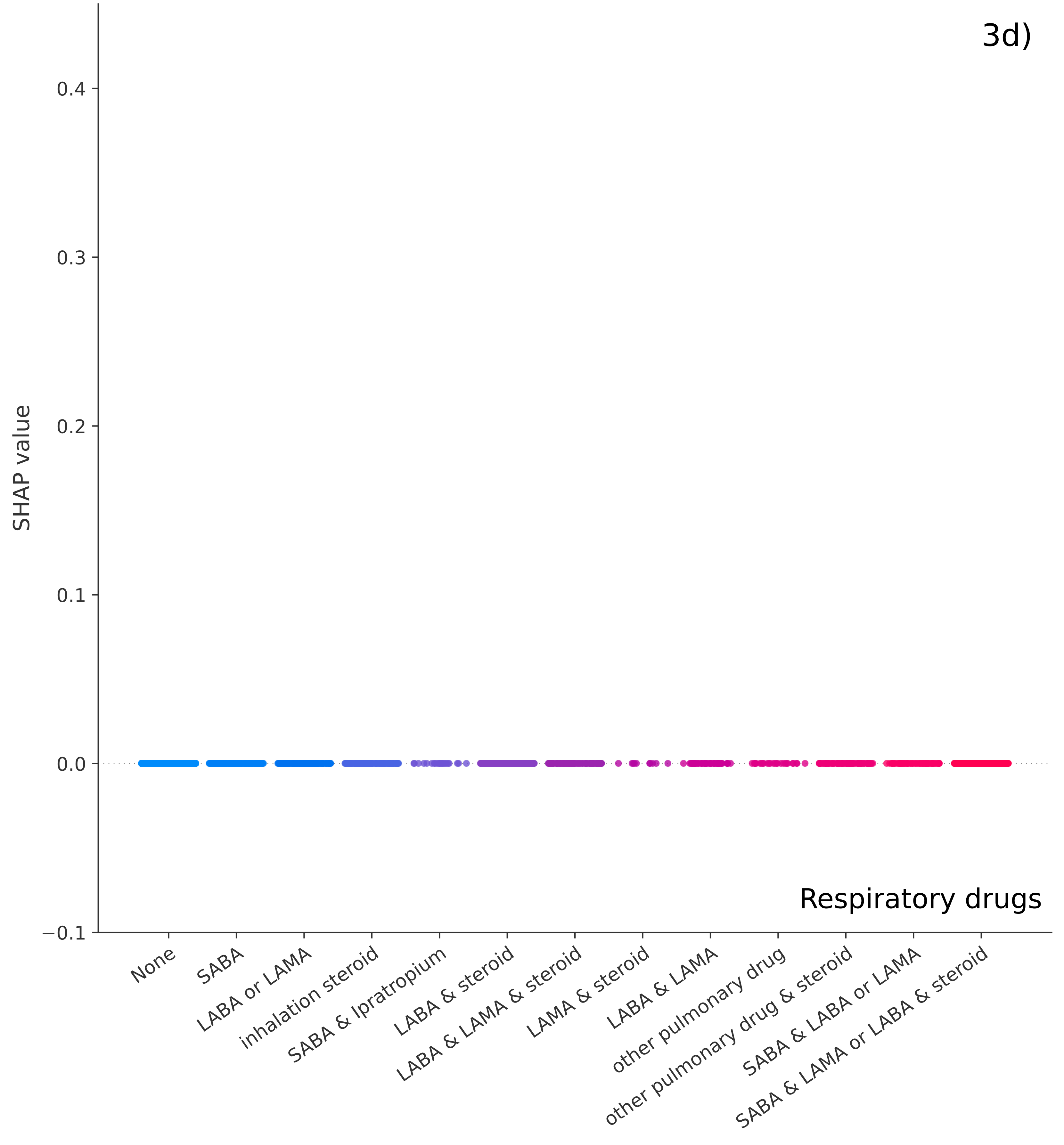
Legend:
3a) Prescribed anticoagulants
VKA: vitamin K antagonists ASA: acetylsalicylic acid DOAC: direct oral anticoagulant ADP: Adenosine diphosphate ACE: angiotensin converting enzyme
3b) Prescribed cardiac drugs
ACE: angiotensin converting enzyme AHT: antihypertensive. Other AHT were defined as AHT 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.



3a)
SHAP value
0.4
0.3
0.2
0.1
0.0
−0.1
Prescribed 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
Age
80
75
70
65
60
55
50


3b)
SHAP value
0.4
0.3
0.2
0.1
0.0
−0.1
Prescribed cardiac drugs
None
diuretics
angiotensin-II/ACE-inhib
Ca^{2+}-antagonists
betablockers
nitrates
other AHT
other IHD-drugs
2 antihypertensives
β-blocker & 1 AHT
3 AHT
β-blocker & 2 AHT
β-blocker & 3 AHT
4 AHT
β-blocker & 4 AHT
other AHT & AHT
nitrates & any AHT
other IHD-drugs & any AHT/nitrate
other antiarrythmics & any ATH
Age
80
75
70
65
60
55
50


3c)
SHAP value
0.4
0.3
0.2
0.1
0.0
−0.1
Psychotropics
None
SSRI/SNRI/NaRI
other AD
antipsychotics
BZ
anticholinergics/memantine
anti-ADHD
NaSSA
other psychotropics
SSRI & other AD
SSRI & NaSSA
SSRI & antipsychotics
SSRI & other psychotropics
BZ+ any psychotropic
antipsychotics + any psychotropic
anti-ADHD + any psychotropic
NaSSA + any psychotropic
other psychotropic & any specified psychotropic
Age
80
75
70
65
60
55
50


3d)
0.4
0.3
0.2
0.1
0.0
−0.1
SHAP value
Respiratory drugs
None
SABA
LABA or LAMA
inhalation steroid
SABA & Ipratropium
LABA & steroid
LABA & LAMA & steroid
LAMA & steroid
LABA & LAMA
other pulmonary drug
other pulmonary drug & steroid
SABA & LABA or LAMA
SABA & LAMA or LABA & steroid