

Sociodemographic determinants of oral anticoagulant prescription in patients with atrial fibrillations: findings from the PINNACLE registry using machine learning



Zahra Azizi, MD, MSc,^{*†1} Andrew T. Ward, PhD,^{†1} Donghyun J. Lee, PhD,[‡] Sanchit S. Gad, BS,[‡] Kanchan Bhasin, MD, MSc,[‡] Robert J. Beetel, PhD,[‡] Tiago Ferreira, MBA,[‡] Sushant Shankar, MS,[‡] John S. Rumsfeld, MD, PhD,[§] Robert A. Harrington, MD,[¶] Salim S. Virani, MD,^{||} Ty J. Gluckman, MD,^{**} Rajesh Dash, MD, PhD,^{*2} Fatima Rodriguez, MD, MPH^{†2}

From the ^{*}Center for Digital Health, Stanford University, Stanford, California, [†]Division of Cardiovascular Medicine and Cardiovascular Institute, Stanford University, Stanford, California, [‡]HealthPals Inc, Millbrae, California, [§]Division of Cardiology, Department of Medicine, University of Colorado School of Medicine, Aurora, Colorado, [¶]Department of Medicine, Stanford University, Stanford, California, ^{||}Sections of Cardiology and Cardiovascular Research, Department of Medicine, Michael E. DeBakey Veterans Affairs Medical Center and Baylor College of Medicine, Houston, Texas, and ^{**}Center for Cardiovascular Analytics, Research and Data Science, Providence Heart Institute, Providence Saint Joseph Health, Portland, Oregon.

BACKGROUND Current risk scores that are solely based on clinical factors have shown modest predictive ability for understanding of factors associated with gaps in real-world prescription of oral anticoagulation (OAC) in patients with atrial fibrillation (AF).

OBJECTIVE In this study, we sought to identify the role of social and geographic determinants, beyond clinical factors associated with variation in OAC prescriptions using a large national registry of ambulatory patients with AF.

METHODS Between January 2017 and June 2018, we identified patients with AF from the American College of Cardiology PINNACLE (Practice Innovation and Clinical Excellence) Registry. We examined associations between patient and site-of-care factors and prescription of OAC across U.S. counties. Several machine learning (ML) methods were used to identify factors associated with OAC prescription.

RESULTS Among 864,339 patients with AF, 586,560 (68%) were prescribed OAC. County OAC prescription rates ranged from 26.8% to 93%, with higher OAC use in the Western United States. Super-

vised ML analysis in predicting likelihood of OAC prescriptions and identified a rank order of patient features associated with OAC prescription. In the ML models, in addition to clinical factors, medication use (aspirin, antihypertensives, antiarrhythmic agents, lipid modifying agents), and age, household income, clinic size, and U.S. region were among the most important predictors of an OAC prescription.

CONCLUSION In a contemporary, national cohort of patients with AF underuse of OAC remains high, with notable geographic variation. Our results demonstrated the role of several important demographic and socioeconomic factors in underutilization of OAC in patients with AF.

KEY WORDS Anticoagulation; Atrial fibrillation; Care gaps; Guidelines; Machine learning

(Heart Rhythm 0² 2023;4:158–168) © 2022 Heart Rhythm Society. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Oral anticoagulation (OAC) reduces the risk of stroke and systemic embolism in patients with atrial fibrillation (AF).

Yet, use of OAC in patients with AF has historically been suboptimal.^{1,2} Previous analyses involving the American College of Cardiology (ACC) PINNACLE (Practice

¹Drs Azizi and Ward are first co-authors. ²Drs Dash and Rodriguez are last co-authors. **Address reprint requests and correspondence:** Dr Fatima Rodriguez, Center for Academic Medicine, Department of Medicine/Division of Cardiovascular Medicine, Mail Code 5687, Stanford University School of Medicine, 453 Quarry Road, Palo Alto, CA 94304. E-mail address: frodrigo@stanford.edu; Twitter: [@FaRodriguezMD](https://twitter.com/FaRodriguezMD).

KEY FINDINGS

- This study complements prior work by leveraging machine learning methods to identify important patient-level predictors of oral anticoagulation prescription care gaps.
- Supervised machine learning analyses outperformed the CHA₂DS₂-VASc (congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, prior stroke or transient ischemic attack or thromboembolism, vascular disease, age 65–74 years, sex category) score at predicting oral anticoagulation use and identified a rank order of associated patient's sociodemographic features beyond clinical factors.
- Significant geographic variation in oral anticoagulation use was observed between counties, with highest rates among patients dwelling in suburban settings and in the Western United States.
- Social determinants of health including household income and clinic size were among highest-ranking features associated with oral anticoagulation prescription.
- The results from this precision population health study should be translated into actionable values for clinicians and care teams to close critical gaps in medical care.

Innovation and Clinical Excellence) Registry from 2008 to 2014 have documented OAC prescription rates ranging between 45% and 61%.^{1,2} Factors contributing to the observed care gaps are numerous and include those at the patient, clinician, and health system levels.

The CHA₂DS₂-VASc score (congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, prior stroke or transient ischemic attack or thromboembolism, vascular disease, age 65–74 years, sex category) has been extensively validated to estimate the risk for the development of stroke based on specific demographic and clinical risk factors and is used to inform treatment decisions about OAC.³ This clinical risk score along with other similar scores have shown modest predicting ability to predict outcomes. One explanation of the modest predictive performance of these scores may be that these scores only include clinical factors and do not consider sociodemographic and geographical variations that are known to be important predictors of cardiovascular outcomes and anticoagulation use.^{4,5} The 2019 American Heart Association/ACC/Heart Rhythm Society guidelines recommend OAC for all AF patients based on a qualifying CHA₂DS₂-VASc score.⁶ Preference is given to direct oral anticoagulants (DOACs) in most patients with AF^{6–8} due to their ease of administration and therapeutic advantages compared with vitamin K antagonists.^{9–13} The current OAC practice patterns for AF patients remain incompletely characterized, with substantial opportunity to

better understand geographic, clinical, and socioeconomic determinants of guideline-directed OAC use.

Machine learning (ML) is a branch of artificial intelligence that leverages data analysis to identify relationships between variables directly from the data. ML encompasses supervised (eg, predicting an outcome) and unsupervised (eg, clustering) methods, which can be used to process complex, high-volume datasets. Such techniques may complement traditional statistical approaches by identifying nonintuitive features or combined patient and site-of-care variables (signatures) to gain insight into patterns and predictors of OAC prescription among patients with AF.

Using the PINNACLE Registry, we sought to describe role of social determinants of health and geographic differences in contemporary OAC prescription practices among patients with AF. We further leveraged a clinical intelligence data platform using ML algorithms to identify predictors of OAC care gaps.

Methods

Data source

We analyzed data from the PINNACLE Registry, which includes 829 practices throughout the United States. Details related to this registry, along with available data elements (eg, patient demographics, comorbidities, vital signs, medications, laboratory values, and recent hospitalizations) have been previously described.^{1,14,15} Waiver of written informed consent and authorization for this study was granted by Chesapeake Research Review Incorporated due to the use of de-identified, retrospective data.

Clinical intelligence engine

For this study, we leveraged CLINT, an analytic engine from the ACC's innovation collaborator, HealthPals Inc (Millbrae, CA). CLINT has a comprehensive array of codified ACC cardiometabolic guidelines that map best practices to individual patient data, allowing for efficient identification of care gaps. Available fields in the PINNACLE Registry (now operated by Veradigm and comprising ~360 structured data elements per patient encounter) were integrated into CLINT and used to derive a CHA₂DS₂-VASc score for each patient. The CHA₂DS₂-VASc score was used to identify care gaps, defined as the percentage of patients who have a Class I indication for OAC according to the 2019 American Heart Association/ACC/Heart Rhythm Society Atrial Fibrillation guideline update.⁶ These evidence-based care gaps were then aggregated into an interactive population dashboard. We also used CLINT to efficiently perform cohort selection and train and evaluate ML models on the PINNACLE data.

Study population

The study population consisted of patients enrolled in the PINNACLE Registry from January 2017 to June 2018. Eligible patients included those with a diagnosis of nonvalvular (or "unspecified") AF at any encounter within

the 18-month survey period and a recorded sex, which is necessary for calculating the CHA₂DS₂-VASc score.

Outcome

The primary outcome was OAC prescription, defined as the presence of at least 1 anticoagulant (apixaban, dabigatran, edoxaban, rivaroxaban, or warfarin) in the most recent 3 months of each patient's record. Prescription rates were calculated by dividing the number of patients prescribed OAC by the total number of OAC-eligible patients with AF. To identify geographic gaps in guideline adherence, we grouped patients with AF into counties based on their clinic's street address. Prescription rates were then calculated for each U.S. county. To ensure sufficient data quality, rates were only calculated for counties with at least 40 patients in the study.

Patient characteristics

Variables were extracted from the PINNACLE registry. For each patient, information was collected from the quarter (3 months) during which the patient's most recent outpatient encounter occurred. In addition to demographic variables, variables representing preselected cardiometabolic comorbidities (dyslipidemia, chronic kidney disease, chronic liver disease, thyroid disease, hemodialysis, prior kidney transplantation, sleep apnea, and stable and unstable angina) and medications prescribed (antihypertensives, antiarrhythmic agents, lipid-modifying therapies, aspirin, antiplatelets other than aspirin [clopidogrel, prasugrel, vorapaxar, ticagrelor], and blood glucose regulation agents) were created by searching for any mention of each within the 3-month period. Laboratory variables (international normalized ratio [INR], glomerular filtration rate, lipid levels) were also collected; if a patient had multiple values from the same lab test in the 3-month period, the values were averaged. Vital signs (heart rate, blood pressure, weight) and insurance information (commercial, Medicaid, Medicare, or other) were included. Clinic information included U.S. census region (South/Midwest/Northeast/West), urbanicity (urban/suburban/rural), and number of patients seen by the clinic. The mean household income was determined using data from the 2016 U.S. Census and the zip code in which the patient's clinic was located. The CHA₂DS₂-VASc score was calculated based on PINNACLE data fields as previously described.¹⁶ Variables were used at the patient level to determine associations and develop models and were aggregated at the county level to explore geographic trends.

ML analysis

ML analyses were completed in Python 3.6 using the Scikit-learn package, version 0.21.2 (Python Software Foundation, Wilmington, DE). In order to identify the most important drivers of guideline-adherent OAC prescriptions, several ML binary classifiers, including logistic regression, LASSO-penalized logistic regression, random forests, and extreme gradient boosting (XGBoost),^{17,18} were trained on

variables derived from the PINNACLE dataset. These classifiers were chosen for their ability to effectively incorporate many variables into the models. The tree-based ML classifiers (random forests and XGBoost) aggregate the predictions of many independent decision trees. This allows these models to capture complex variable interactions while simultaneously minimizing variance.

As a baseline comparison, we analyzed the ability of the CHA₂DS₂-VASc score to directly predict which patients would receive an OAC prescription.¹ To compare directly with the CHA₂DS₂-VASc score, we trained the 4 ML models to predict OAC prescription using only the variables comprising the CHA₂DS₂-VASc score: sex, age, heart failure, hypertension, diabetes, peripheral artery disease, peripheral vascular disease, prior myocardial infarction, coronary artery bypass grafting surgery, percutaneous coronary intervention, ischemic stroke, and transient ischemic attack.

As ML models (particularly tree-based models) are able to effectively use a large number of variables that can be correlated with each other, we next included a wide range of clinical, demographic, and geographic variables in "enhanced" variants of the ML models. In addition to the CHA₂DS₂-VASc variables, clinic information (region, urbanicity, number of patients), demographic factors (race/ethnicity, mean household income), medical comorbidities, medications prescribed, vital signs, laboratory data, and insurance information were used to train and evaluate the enhanced ML classifiers. For variables with continuous values, missing fields were imputed using the median value of that variable across the dataset.

Data were split into training (80%) and testing (20%) sets. Within the training set, 5-fold cross-validation was used to tune hyperparameters, such as regularization parameters, maximum tree depth, and number of trees. Hyperparameters were tuned to control the models' complexities and prevent overfitting. Models were compared based on the area under the receiver-operating characteristic curve (AUROC), also known as the C-statistic. Once the best hyperparameters were selected for each ML classifier, models with these hyperparameters were retrained on the entire training set. The final AUROCs, model accuracy, precision, recall, and area under the precision-recall curve of both regular and "enhanced" ML models, as well as the CHA₂DS₂-VASc score, were reported on the testing set.

We also assessed the feature importance of the model with the highest testing set AUROC in order to understand how much weight the model places on each of the expanded set of covariates in determining OAC prescription probability. This analysis was only performed on the testing set. Traditionally, feature importance for random forests is reported using the decrease in the Gini impurity in the training process, but this has a number of shortcomings,¹⁹ particularly in the presence of correlated variables and variables of mixed types (binary/categorical/continuous). As such, we ranked variables using permutation importance,²⁰ a method that randomly permutes values in columns of the test data and measures decrease in performance. Notably, while the permutation importance represents the overall magnitude of

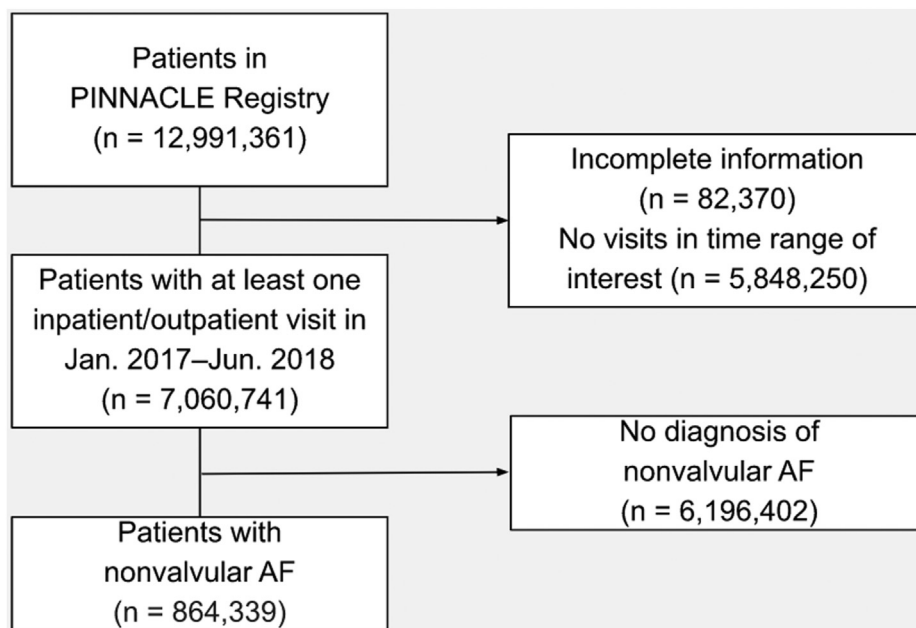


Figure 1 Cohort diagram. Flow chart detailing the inclusion and exclusion criteria used for identifying this study population from the PINNACLE (Practice Innovation and Clinical Excellence) dataset. AF = atrial fibrillation.

influence for each feature on the OAC prescription rate, the polarity of influence of any given feature (eg, CHA₂DS₂-VASc score) may be a combination of positive and negative statistical associations that may depend on the numerical value of the feature itself or other variable inputs. To offer insight into the polarity of variable influence, we plot OAC rates against the most important variables ([Supplemental Figure 1](#)).

Data availability

We declare that the data supporting the findings of this study are available within the article and its Supplemental information files.

The raw data that support the findings of this study are also available from National Cardiovascular Data Registry PINNACLE Registry, but restrictions apply to the availability of these data, which were used under license for the current study, and so they are not publicly available. Data are, however, available from the authors upon reasonable request and with permission of the ACC.

The linked data used in this analysis were deidentified, so the study was exempt from the requirement for review board approval and . The research reported in this article adhered to Helsinki Declaration guidelines for human research.

Results

Descriptive patterns in OAC prescription

Between January 1, 2017, and June 30, 2018, there were 864,339 patients with AF in the registry ([Figure 1](#)). [Table 1](#) shows patient-level characteristics for the study population

by OAC use. A total of 586,554 (68%) received OAC, of which, 69% (n = 401,953) were prescribed a DOAC. Most of the AF patients (85% [n = 734,288]) met contemporary Class I indications for an OAC prescription; approximately 70% of these patients (n = 520,909) were prescribed OAC.

Patients who were prescribed OAC were more likely to reside in the Western United States and in suburban counties, be visited in greater clinic size, be non-Hispanic White, be older, have greater household income, and be insured through Medicare.

Moreover, these patients were more likely to have greater body mass index; have a history of hypertension, heart failure, stroke, diabetes, chronic kidney disease, or sleep apnea; and be treated with antihypertensive, antiarrhythmic, lipid-modifying, or blood glucose–regulating medications.

Several continuous variables including weight were not available for over 50% of patients, for which imputation methods were employed as described previously; completeness information for continuous variables is available in [Supplemental Table 1](#). OAC prescription rates increased with an increasing CHA₂DS₂-VASc score from 0 to 4, with a slight decrease among those with CHA₂DS₂-VASc scores >4 ([Supplemental Figure 1](#)).

Geographic patterns of OAC use

[Figure 2](#) displays the geographic patterns of OAC prescription rates by county. Wide variation was observed, with county-level OAC prescription rates ranging from 26.8% to 93.2%. The counties with the lowest OAC prescription rates (<60% OAC coverage) tended to be in urban areas and were more common in Iowa, Florida, Louisiana, Texas, and

Table 1 Characteristics of AF patients

	Total (N = 864,339)	OAC Prescription (n = 586,554)	No OAC Prescription (n = 277,785)	P Value
Mean household income (×\$1,000)	69.43 ± 13.15	69.50 ± 13.12	69.28 ± 13.21	<.001
Age, y	73.54 ± 11.43	74.48 ± 10.38	71.57 ± 13.15	<.001
Male	490,102 (57)	332,949 (57)	157,153 (57)	.096
Weight, kg	88.47 ± 24.38	89.52 ± 24.68	86.23 ± 23.56	<.001
CHA ₂ DS ₂ -VASc score	3.60 ± 1.71	3.74 ± 1.64	3.30 ± 1.81	<.001
Eligible for OAC*	734,288 (85)	520,909 (89)	213,379 (77)	<.001
Race/ethnicity				
Hispanic	25,532 (3)	16,834 (2.9)	8,698 (3.1)	<.001
Non-Hispanic White	572,814 (66)	393,402 (67)	179,412 (65)	
Non-Hispanic Black	37,833 (4.4)	24,671 (4.2)	13,162 (4.7)	
Other	13,268 (1.5)	8,864 (1.5)	4,404 (1.6)	
Missing	214,892 (25)	142,783 (24)	72,109 (26)	
Clinic location				
West	161,060 (19)	111,685 (19)	49,375 (18)	<.001
Northeast	182,001 (21)	124,672 (21)	57,329 (21)	
Midwest	131,658 (15)	89,046 (15)	42,612 (15)	
South	436,030 (50)	292,118 (50)	143,912 (52)	
Urban	182,659 (21)	123,413 (21)	59,246 (21)	
Suburban	136,431 (16)	96,446 (16)	39,985 (14)	
Rural	32,420 (3.8)	22,109 (3.8)	10,311 (3.7)	
Size of clinic (number of patients)	19,705.51 ± 16,583.81	19,994.35 ± 16,805.25	19,095.63 ± 16,089.24	<.001
Insurance type				
Private	500,882 (58)	340,187 (58)	160,695 (58)	<.001
Medicaid	54,517 (6.3)	35,652 (6.1)	18,865 (6.8)	
Medicare	526,572 (61)	375,615 (64)	150,957 (54)	
State	7,839 (0.91)	5,155 (0.88)	2,684 (0.97)	
Other	35,070 (4.1)	24,250 (4.1)	10,820 (3.9)	
None	3,225 (0.37)	2,147 (0.37)	1,078 (0.39)	
Clinic and lab values				
Heart rate, beats/min	72.20 ± 13.51	72.60 ± 13.72	71.41 ± 13.03	<.001
Systolic BP, mm Hg	127.97 ± 17.20	127.77 ± 17.12	128.40 ± 17.37	<.001
Diastolic BP, mm Hg	73.57 ± 10.40	73.45 ± 10.34	73.82 ± 10.52	<.001
Total cholesterol, mg/dL	158.82 ± 40.99	156.63 ± 39.95	163.63 ± 42.81	<.001
HDL cholesterol, mg/dL	50.13 ± 16.81	49.78 ± 16.61	50.90 ± 17.22	<.001
LDL cholesterol, mg/dL	86.83 ± 34.65	85.40 ± 34.03	90.00 ± 35.77	<.001
Triglyceride, mg/dL	125.49 ± 70.64	125.13 ± 69.96	126.27 ± 72.10	.008
INR	2.14 ± 2.10	2.23 ± 1.97	1.75 ± 2.54	<.001
GFR, mL/min/1.73 m ²	63.67 ± 22.96	62.74 ± 22.16	66.11 ± 24.76	<.001
LVEF, %	54.98 ± 12.94	54.34 ± 13.18	56.48 ± 12.23	<.001
Comorbidities				
Hypertension	664,713 (77)	463,872 (79)	200,841 (72)	<.001
Dyslipidemia	517,586 (60)	360,239 (61)	157,347 (57)	<.001
Heart failure	238,781 (28)	177,488 (30)	61,293 (22)	<.001
Stable angina	87,749 (10)	56,458 (9.6)	31,291 (11)	<.001
Unstable angina	26,105 (3)	15,994 (2.7)	10,111 (3.6)	<.001
Transient ischemic attack	56,425 (6.5)	40,649 (6.9)	15,776 (5.7)	<.001
Ischemic stroke	67,588 (7.8)	48,479 (8.3)	19,109 (6.9)	<.001
Coronary artery disease	375,678 (43)	257,035 (44)	118,643 (43)	<.001
Myocardial infarction	55,161 (6.4)	35,313 (6)	19,848 (7.1)	<.001
Peripheral artery disease	99,234 (11)	68,325 (12)	30,909 (11)	<.001
Peripheral vascular disease	74,430 (8.6)	51,771 (8.8)	22,659 (8.2)	<.001
Coronary artery bypass grafting	63,543 (7.4)	40,949 (7)	22,594 (8.1)	<.001
Percutaneous coronary intervention	76,282 (8.8)	49,994 (8.5)	26,288 (9.5)	<.001
Type 2 diabetes	217,996 (25)	157,247 (27)	60,749 (22)	<.001
Chronic kidney disease	89,262 (10)	63,272 (11)	25,990 (9.4)	<.001
Chronic liver disease	90,557 (10)	63,957 (11)	26,600 (9.6)	<.001
Hemodialysis	2,370 (0.27)	1,492 (0.25)	878 (0.32)	<.001
Kidney transplant	590 (0.068)	404 (0.069)	186 (0.067)	.783
Hyperthyroidism	8,024 (0.93)	5,511 (0.94)	2,513 (0.9)	.117
Hypothyroidism	55,172 (6.4)	37,933 (6.5)	17,239 (6.2)	<.001
Sleep apnea	82,975 (9.6)	60,563 (10)	22,412 (8.1)	<.001
Medications				

Table 1 (Continued)

	Total (N = 864,339)	OAC Prescription (n = 586,554)	No OAC Prescription (n = 277,785)	P Value
Antiplatelets	487,204 (56)	291,608 (50)	195,596 (70)	<.001
Antiplatelets (without aspirin)	105,056 (12)	64,248 (11)	40,808 (15)	<.001
Antiarrhythmic agents	327,998 (38)	253,181 (43)	74,817 (27)	<.001
Lipid-modifying agents	533,753 (62)	390,012 (66)	143,741 (52)	<.001
Blood glucose regulation agents	167,735 (19)	126,588 (22)	41,147 (15)	<.001
Antihypertensives	775,798 (90)	553,571 (94)	222,227 (80)	<.001
Aspirin	464,821 (54)	276,320 (47)	188,501 (68)	<.001
Prasugrel	6409 (0.74)	4232 (0.72)	2177 (0.78)	.002
Ticagrelor	7104 (0.82)	4249 (0.72)	2855 (1)	<.001
Clopidogrel	97,316 (11)	59,874 (10)	37,442 (13)	<.001
Vorapaxar	78 (0.009)	34 (0.0058)	44 (0.016)	<.001
Anticoagulants				
Warfarin	232,538 (27)	232,538 (40)	—	<.001
Apixaban	232,720 (27)	232,720 (40)	—	<.001
Dabigatran	54,136 (6.3)	54,136 (9.2)	—	<.001
Rivaroxaban	147,594 (17)	147,594 (25)	—	<.001
Edoxaban	4291 (0.5)	4291 (0.73)	—	<.001
DOACs	401,953 (47)	401,953 (69)	—	<.001

Values are mean \pm SD or n (%). Patients were stratified by whether they received an OAC prescription.

AF = atrial fibrillation; BP = blood pressure; CHA₂DS₂-VASc = congestive heart failure, hypertension, age \geq 75 years, diabetes mellitus, prior stroke or transient ischemic attack or thromboembolism, vascular disease, age 65–74 years, sex category; DOAC = direct oral anticoagulant; GFR = glomerular filtration rate; HDL = high-density lipoprotein; INR = international normalized ratio; LDL = low-density lipoprotein; LVEF = left ventricular ejection fraction; OAC = oral anticoagulation.

*Class I indication for OAC was determined by whether a patient had an elevated CHA₂DS₂-VASc score as specified by contemporary guidelines (CHA₂DS₂-VASc score \geq 2 for men, \geq 3 for women).⁴

Virginia. In contrast, nearly all counties in Arizona, Oklahoma, Connecticut, Vermont, and Maine had OAC prescription rates above 60%. Patient characteristics by each quartile of OAC prescription rates are further detailed in [Supplemental Table 2](#).

ML insights: Determinants of OAC prescription

The enhanced XGBoost model performed best in its ability to identify whether patients did or did not receive OAC with a test AUROC of 0.811 (95% confidence interval 0.809–0.813). This significantly surpassed the predictive performance of the CHA₂DS₂-VASc score (AUROC 0.571, 95% confidence interval 0.569–0.574) ([Figure 3](#)). Every enhanced ML model outperformed all versions of the ML models, which only relied on the CHA₂DS₂-VASc variables ([Table 2](#)). [Figure 4](#) shows features in order of permutation importance within the enhanced XGBoost model. The most predictive patient features include (1) use of aspirin, antihypertensives, antiarrhythmic agents, lipid-modifying agents, or antiplatelets; (2) age; (3) mean household income; (4) INR values; (5) clinic size; (6) patient weight; and (7) U.S. region. Beyond age, variables included in the CHA₂DS₂-VASc score had low importance in the enhanced random forest model (ranking 12th, 21st, 24th, 25th, 30th, and lower). [Supplemental Figure 1](#) displays the different positive and negative associations of each feature in more detail.

Discussion

In a contemporary cohort of U.S. patients with AF, 68% were treated with OAC. Of the remaining third of AF patients not

on OAC, most met a Class I indication for OAC use by contemporary guidelines. Significant geographic variation in OAC use was observed between counties, with highest rates among patients dwelling in suburban settings and in the Western United States. Supervised ML analyses outperformed the CHA₂DS₂-VASc score at predicting OAC use and identified a rank order of associated patient's sociodemographic features beyond clinical factors. The strongest associations were the use of aspirin, antihypertensives, antiarrhythmic agents, lipid-modifying agents, and INR values, as well as the features of age, mean household income, clinic size, patient weight, and geographic region. Our results are largely consistent with prior findings of disparities in OAC prescription by patient characteristics, site of care, and geographic region. One such analysis²¹ found that, compared with those prescribed OAC, patients with AF prescribed aspirin as their sole antithrombotic therapy were more often located in the South and West, in nonurban settings, and in practices with larger patient volumes.

While other studies have demonstrated an association between greater burden of clinical comorbidities²² and lower likelihood of OAC,²² sleep apnea has been associated with increased OAC use.²³ Our analysis revealed that the largest contributions to the predictive model of OAC use were prescriptions related to other comorbid conditions, including hypertension, congestive heart failure, diabetes, and stroke. Similar to other analyses,²³ we also found that use of aspirin and other forms of antiplatelet therapy was associated with lower rates of OAC prescriptions.²⁴ This is possibly due to the increased risk of bleeding in these patients. OAC use was greater in patients who were on antiarrhythmic agents,

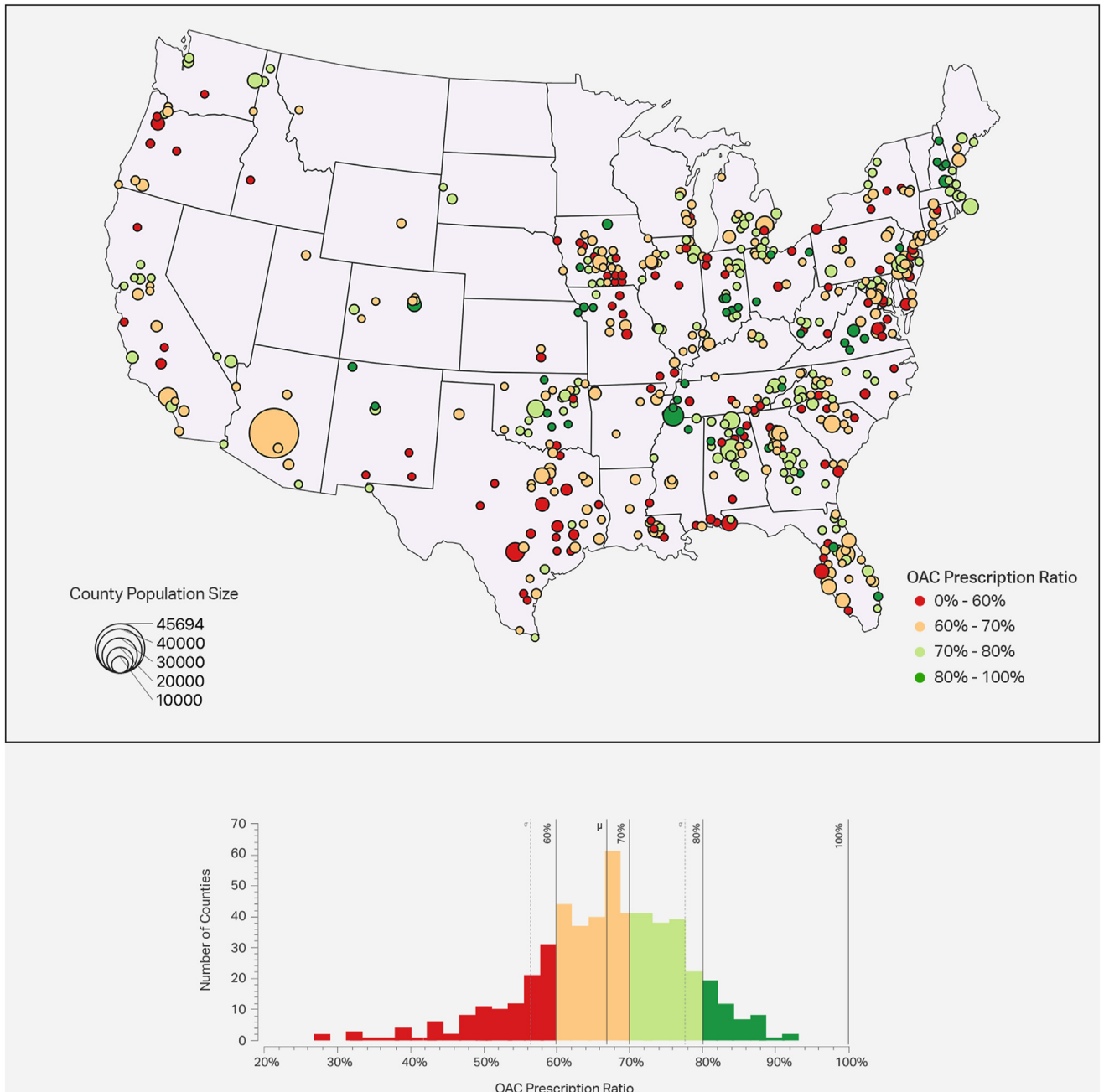


Figure 2 Oral anticoagulation (OAC) prescription rates by county. OAC prescription rates are shown at a population level, split into 4 quartiles. Prescription rates are defined as [atrial fibrillation patients with OAC prescriptions] / [atrial fibrillation patients]. The circle size denotes the number of PINNACLE (Practice Innovation and Clinical Excellence) patients treated within that area. The OAC prescription rates are shown by color, with red/orange areas indicating worse rates and green areas indicating better rates. A histogram is also shown, indicating how many counties fall in each bin of OAC adherence.

which may be explained by the increased recurrence and severity of AF in patients on antiarrhythmic therapy.

Our results also demonstrated geographical variation in OAC prescription, in which the counties with the lowest OAC prescription rates (<60% OAC coverage) tended to be in urban areas and were more common in Iowa, Florida, Louisiana, Texas, and Virginia. In contrast, nearly all counties in Arizona, Oklahoma, Connecticut, Vermont, and Maine had OAC prescription rates above 60%. Similar

results were obtained in the study by Hernandez and colleagues,⁵ who reported large geographical variations in use of OAC for stroke prevention in patients with AF. In this study, the Midwest and Northwest had a higher likelihood of OAC initiation compared with the South, which had the lowest likelihood of OAC use and a higher risk of stroke.

One important finding of our analysis was the role of social determinants of health, including household income

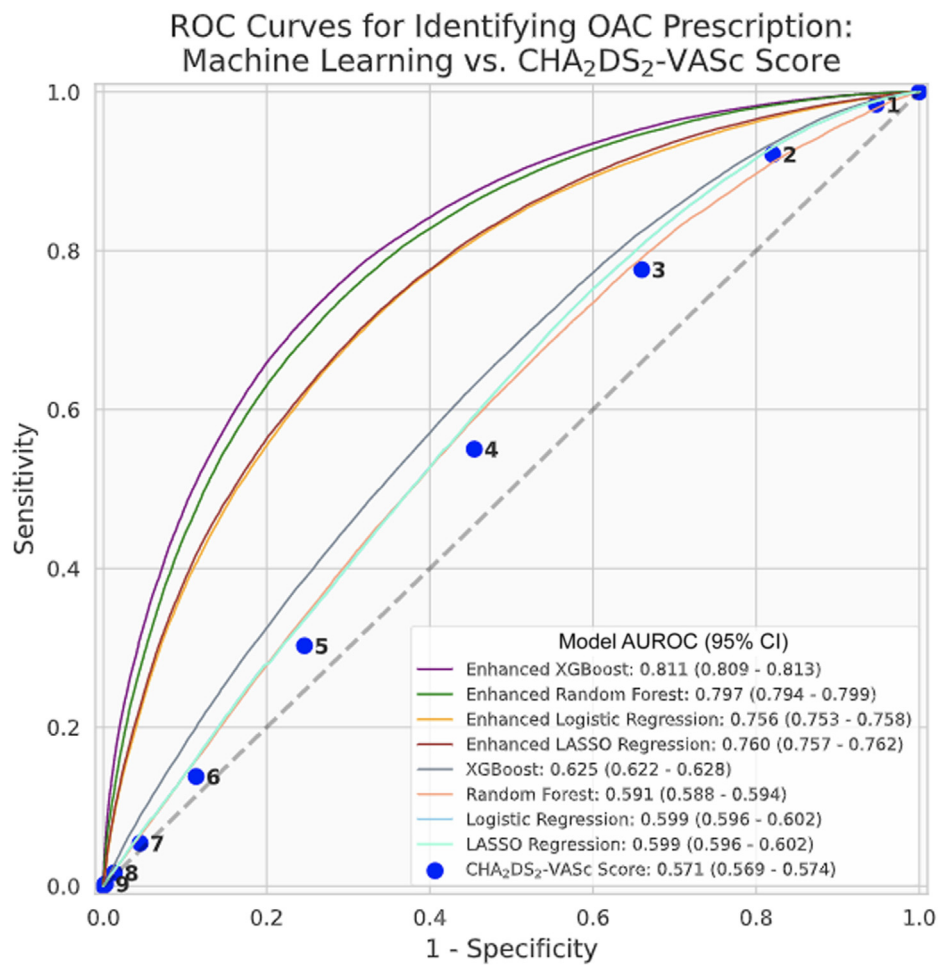


Figure 3 Receiver-operating characteristic (ROC) curves for identifying oral anticoagulation (OAC) prescription: machine learning (ML) vs CHA₂DS₂-VASc (congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, prior stroke or transient ischemic attack or thromboembolism, vascular disease, age 65–74 years, sex category) score. ROC curves for (1) the CHA₂DS₂-VASc score, (2) 4 ML models (XGBoost, random forest, logistic regression, LASSO regression) using only covariates considered in the CHA₂DS₂-VASc score, and (3) 4 “enhanced” ML models that were trained on additional clinical comorbidities, medication usage, vital signs, laboratory data, insurance information, and socio- and geodemographic variables. Metrics were calculated on held-out test data. AUROC = area under the receiver operating characteristic curve; CI = confidence interval.

and clinic size in OAC prescription and adherence. Cost of medication and follow-up of OAC can influence the prescription rates.^{25,26} Unequal access to OAC in socioeconomically

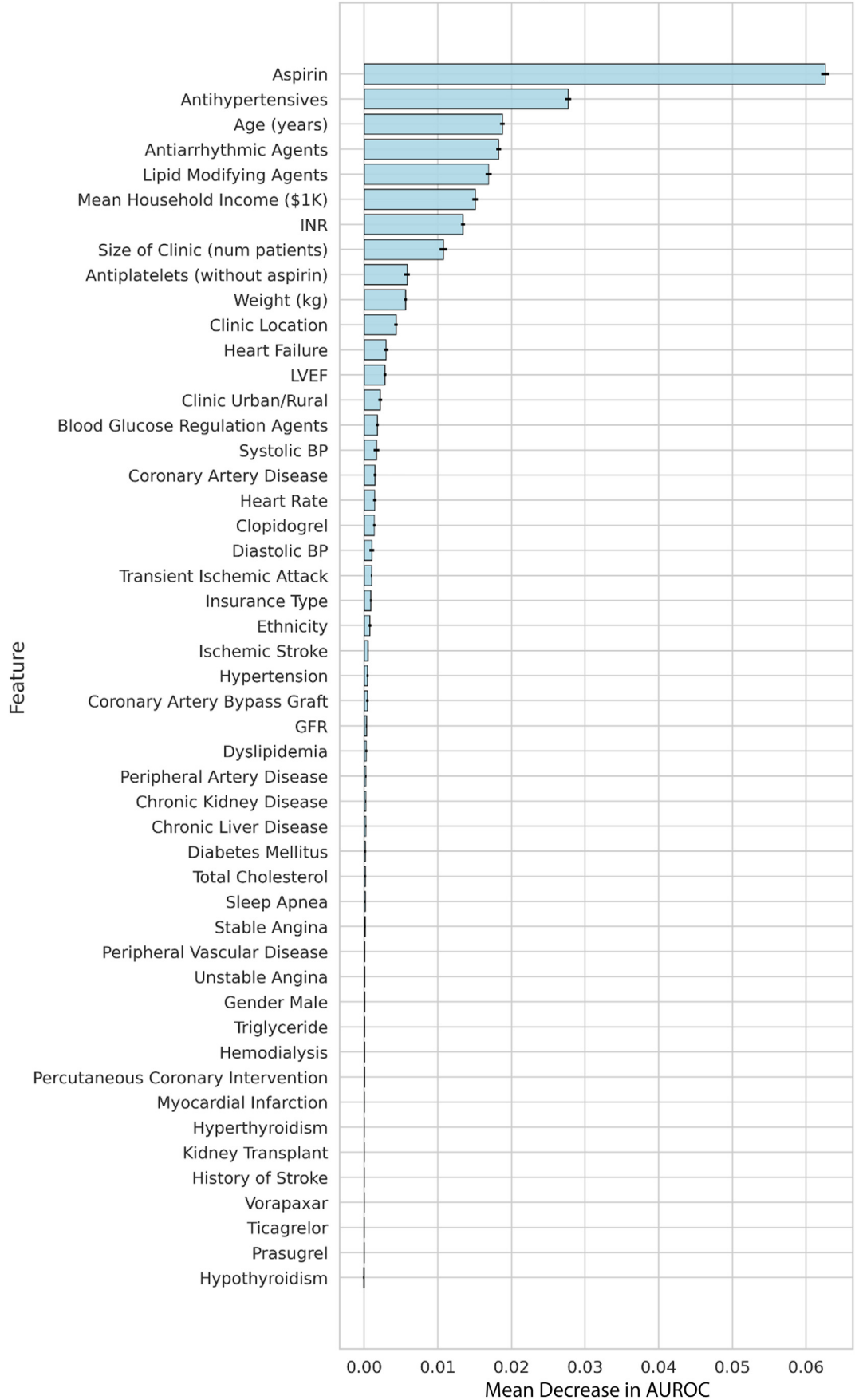
disadvantaged patients and different geographical areas have been shown in previous studies.²⁶ In a study by Llorca and colleagues,²⁵ those living in more socioeconomically

Table 2 Summary of model performances for predicting oral anticoagulation prescription in training (5-fold cross-validation) and test sets

		Regular model: CHA ₂ DS ₂ -VASc components					Enhanced ML model: CHA ₂ DS ₂ -VASc components + new features				
		Accuracy	AUROC	PRAUC	Precision	Recall	Accuracy	AUROC	PRAUC	Precision	Recall
XGBOOST	Test set	0.69	0.62	0.76	0.70	0.96	0.77	0.81	0.89	0.79	0.89
	Cross-validation	0.70	0.64	0.77	0.70	0.95	0.78	0.83	0.90	0.80	0.89
Logistic regression	Test set	0.69	0.60	0.73	0.70	0.96	0.73	0.75	0.86	0.77	0.87
	Cross-validation	0.69	0.60	0.73	0.70	0.96	0.74	0.76	0.86	0.76	0.88
Random forest	Test set	0.68	0.59	0.73	0.68	0.99	0.76	0.79	0.85	0.79	0.88
	Cross-validation	0.75	0.78	0.88	0.76	0.93	0.99	0.99	0.85	0.99	0.99
LASSO-penalized logistic regression	Test set	0.69	0.60	0.73	0.70	0.96	0.74	0.76	0.88	0.76	0.89
	Cross-validation	0.69	0.60	0.73	0.70	0.96	0.74	0.76	0.99	0.76	0.89

AUROC = area under the receiver-operating characteristic curve; CHA₂DS₂-VASc = congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, prior stroke or transient ischemic attack or thromboembolism, vascular disease, age 65–74 years, sex category; ML = machine learning; PRAUC = area under the precision-recall curve.

Permutation Feature Importances for Enhanced XGBoost



deprived and rural areas had lower OAC prescription rates. Moreover, previous studies by Essien and colleagues^{27,28} showed lower initiation of OAC for Black patients and lower DOAC use for Black and Hispanic patients. This was also evident in our descriptive results; however, these did not emerge as high-ranking features in our ML models, which may be due to the low sample sizes of these populations in our database. Therefore, there is a need to condition prescription patterns by sociodemographic factors besides clinical risk factors.

In the absence of other prediction models to estimate OAC use, we utilized the CHA₂DS₂-VASc score. Even though the CHA₂DS₂-VASc score was initially designed to predict thromboembolic risk, prior work has demonstrated increased odds of OAC prescribing with increasing CHA₂DS₂-VASc score.¹ As such, we hypothesized that it would be modestly predictive of OAC use. We instead found that the performance of CHA₂DS₂-VASc score was only slightly above chance. Furthermore, the enhanced ML models (which incorporated additional social, geographic, and clinical variables) significantly outperformed the ML models that were limited to risk factors in the CHA₂DS₂-VASc score. This finding suggests that a range of social and clinical determinants of health likely underlie much of the observed variation in OAC guideline adherence.²⁴

Our study extends and complements prior work by leveraging ML methods to identify important patient-level predictors of OAC prescriptions. Exploring the variables selected by ML adds unique insights beyond what traditional regression analysis provides. In real-world datasets, a number of variables may be unavailable for large numbers of patients. When these variables are present, they display markedly nonlinear, and even nonmonotonic trends. Additive models such as the CHA₂DS₂-VASc score, which assume equal weights for all risk factors, are unable to fully capture these associations. Furthermore, the rank order of a given feature's influence on OAC prescribing, considering all other possible permutations of other concomitant features, would not be uncovered by less sophisticated models. While traditional multivariate approaches, such as Bayesian hierarchical linear models, are adept at identifying independent associations between a given patient feature and an outcome, ML enables the integration of potentially hundreds of different features, with varying levels of missingness, to determine the collective associations with clinical outcomes. These observations on an established, longitudinal patient registry demonstrate that ML offers unique additive value—and should continue to be leveraged—to identify nonlinear associations between patient features and clinical management practices.

Clinical implications

It is important to translate these findings into actionable value for clinicians and care teams to close critical gaps in medical care. One proposed approach is to utilize an analytics platform to apply guideline-driven insights, both longitudinally and in real time, for every patient record within a health system at once. In the case of OAC use in eligible AF patients, this precision population health engine may alert care teams to focus their attention on specific patient cohorts with confirmed care gaps for OAC, or, upon patient cohorts who are at the highest risk of developing an OAC care gap, as may be the case with the predictive model presented here.

This patient-centered novel approach will provide an accurate tool for clinical decision making not only by incorporating clinical factors considered in the previous risk scores, but also by including social determinants of health and geographical variations for risk profiling of patients with AF.

Limitations

Our results should be interpreted in the context of several limitations. Because our data included patients enrolled predominantly within outpatient cardiology practices, OAC prescribing patterns may not be generalized to noncardiology practices. Incomplete or missing data may also have impacted our findings. For example, if a feature was not reported in the electronic health record for a patient (eg, a history of stroke), it was interpreted in this analysis as the absence of stroke in the calculation of that patient's CHA₂DS₂-VASc score. Patients had differing numbers of recorded encounters in the data, which led to differing levels of data completeness. Because this is a voluntary registry, sites that participate in the PINNACLE Registry may not be nationally representative, and some regions are not well represented. It is likely that the noted sociodemographic disparities may be greater in other non-registry-participating sites. Unlike a randomized controlled trial, the inference of causation is not possible due to the many uncontrolled factors not recorded.¹ There are other confounding factors related to underprescription of OAC including utilization of left atrial appendage occlusion devices, which were not captured in our study. Moreover, increased risk of bleeding is an important factor associated with lower prescription rates. Future studies assessing this risk using related clinical risk scores (HAS-BLED [hypertension, abnormal renal or liver function, stroke, bleeding, labile international normalized ratio, elderly, drugs or alcohol]) are warranted. As this study was based on data in 2017 to 2018, there may be lower-than-expected DOAC use or higher-than-expected acetylsalicylic

Figure 4 Predictive importance of individual clinical features on the likelihood of being prescribed an oral anticoagulation. The highest-performing machine learning model, XGBoost, was used to determine the rank order of features associated with oral anticoagulation prescriptions. This feature importance was measured using the permutation importance metric. With the fully trained model, independent variables were randomly shuffled, removing the relationships learned by the machine learning model, and the decrease in model performance was assessed. The average decrease in performance across 5 independent runs, and the standard deviation of those runs (black error bars) is shown for each variable. AUROC = area under the receiver operating characteristic curve; BP = blood pressure; GFR = glomerular filtration rate; INR = international normalized ratio; LVEF = left ventricular ejection fraction.

acid use for a CHA₂DS₂-VASc = 1 population, given that this would be based on the 2014 guidelines. Future work will explore the specific determinants of DOAC usage vs warfarin usage in the OAC groups, as current guidelines recommend DOACs for most AF patients. Future investigations will also aim to transform these insights into a population health-focused strategy to better target evidence-based interventions that promote closure of gaps in care.

Conclusions

In a contemporary national cohort of patients with AF, almost a third of patients with AF failed to receive OAC, with significant geographic practice variations. Specific, ML-derived predictors of OAC prescription were identified and offer complementary information to traditional analytic methods. Our results demonstrated the role of several important demographic and socioeconomic factors in underutilization of OAC in patients with AF. Therefore, by combining large, representative real-world datasets with ML techniques, features beyond clinical factors contributing to OAC underuse may be identified to inform targets for quality improvement.

Funding Sources: Fatima Rodriguez received support from the National Heart, Lung, and Blood Institute of the National Institutes of Health (1K01HL144607), the American Heart Association/Robert Wood Johnson Harold Amos Medical Faculty Development Program, and Grant #2022051 from the Doris Duke Charitable Foundation. The other authors have no funding sources to disclose.

Disclosures: Fatima Rodriguez serves as an advisor to HealthPals and has served as a consultant for Novartis and Novo Nordisk. Rajesh Dash is a consultant to HealthPals and Bayer. John S. Rumsfeld is Chief Innovation and Chief Science Officer for the American College of Cardiology. Andrew T. Ward, Donghyun J. Lee, Sanchit S. Gad, Kanchan Bhasin, Robert J. Beutel, Tiago Ferreira and Sushant Shankar, are paid employees of and report equity from HealthPals. The remaining authors have no competing interests to declare.

Authorship: All authors attest they meet the current ICMJE criteria for authorship.

Patient Consent: Waiver of written informed consent and authorization for this study was granted by Chesapeake Research Review Incorporated due to the use of de-identified, retrospective data.

Ethics Statement: The research reported in this paper adhered to Helsinki Declaration guidelines for human research.

References

- Hsu JC, Maddox TM, Kennedy KF, et al. Oral anticoagulant therapy prescription in patients with atrial fibrillation across the spectrum of stroke risk: insights from the NCDR PINNACLE Registry. *JAMA Cardiol* 2016;1:55–62.
- Marzec LN, Wang J, Shah ND, et al. Influence of direct oral anticoagulants on rates of oral anticoagulation for atrial fibrillation. *J Am Coll Cardiol* 2017;69:2475–2484.
- Lip GY, Nieuwlaat R, Pisters R, Lane DA, Crijns HJ. Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest* 2010;137:263–272.
- Havranek EP, Mujahid MS, Barr DA, et al. Social determinants of risk and outcomes for cardiovascular disease. *Circulation* 2015;132:873–898.
- Hernandez I, Saba S, Zhang Y. Geographic variation in the use of oral anticoagulation therapy in stroke prevention in atrial fibrillation. *Stroke* 2017;48:2289–2291.
- January CT, Wann LS, Calkins H, et al. 2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Rhythm Society. *J Am Coll Cardiol* 2019;74:104–132.
- Yao X, Abraham NS, Alexander GC, et al. Effect of adherence to oral anticoagulants on risk of stroke and major bleeding among patients with atrial fibrillation. *J Am Heart Assoc* 2016;5:e003074.
- Jackevicius CA, Tsadok MA, Essebag V, et al. Early non-persistence with dabigatran and rivaroxaban in patients with atrial fibrillation. *Heart* 2017;103:1331–1338.
- Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med* 2011;365:883–891.
- Granger CB, Alexander JH, McMurray JJV, et al. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011;365:981–992.
- Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009;361:1139–1151.
- Giugliano RP, Ruff CT, Braunwald E, et al. Edoxaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2013;369:2093–2104.
- January CT, Wann LS, Alpert JS, et al. 2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: executive summary: a report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the Heart Rhythm Society. *Circulation* 2014;130:2071–2104.
- Chan PS, Maddox TM, Tang F, Spinler S, Spertus JA. Practice-level variation in warfarin use among outpatients with atrial fibrillation (from the NCDR PINNACLE program). *Am J Cardiol* 2011;108:1136–1140.
- Messenger JC, Ho KKL, Young CH, et al. The National Cardiovascular Data Registry (NCDR) Data Quality Brief: the NCDR Data Quality Program in 2012. *J Am Coll Cardiol* 2012;60:1484–1488.
- Thompson LE, Maddox TM, Lei L, et al. Sex differences in the use of oral anticoagulants for atrial fibrillation: a report from the National Cardiovascular Data Registry (NCDR) PINNACLE Registry. *J Am Heart Assoc* 2017;6:e005801.
- Breiman L. Random Forests. *Mach Learn* 2011;45:5–32.
- Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY: ACM; 2016. p. 785–794.
- Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics* 2007;8:25.
- Altmann A, Toloşi L, Sander O, Lengauer T. Permutation importance: a corrected feature importance measure. *Bioinformatics* 2010;26:1340–1347.
- Hsu JC, Maddox TM, Kennedy K, et al. Aspirin instead of oral anticoagulant prescription in atrial fibrillation patients at risk for stroke. *J Am Coll Cardiol* 2016;67:2913–2923.
- Savarese G, Sartipy U, Friberg L, Dahlström U, Lund LH. Reasons for and consequences of oral anticoagulant underuse in atrial fibrillation with heart failure. *Heart* 2018;104:1093–1100.
- Johnson KG, Johnson DC. Obstructive sleep apnea is a risk factor for stroke and atrial fibrillation. *Chest* 2010;138:239. ; author reply 239–240.
- Lubitz SA, Khurshid S, Weng L-C, et al. Predictors of oral anticoagulant non-prescription in patients with atrial fibrillation and elevated stroke risk. *Am Heart J* 2018;200:24–31.
- Dalmáu Llorca MR, Aguilar Martín C, Carrasco-Querol N, et al. Gender and socioeconomic inequality in the Prescription of Direct Oral Anticoagulants in Patients with Non-Valvular Atrial Fibrillation in Primary Care in Catalonia (Fantas-TIC Study). *Int J Environ Res Public Health* 2021;18:10993.
- Teppo K, Jaakkola J, Biancarfi F, et al. Association of income and educational levels with adherence to direct oral anticoagulant therapy in patients with incident atrial fibrillation: a Finnish nationwide cohort study. *Pharmacol Res Perspect* 2022;10:e00961.
- Essien UR, Kim N, Magnani JW, et al. Association of race and ethnicity and anticoagulation in patients with atrial fibrillation dually enrolled in Veterans Health Administration and Medicare: effects of Medicare Part D on prescribing disparities. *Circ Cardiovasc Qual Outcomes* 2022;15:e008389.
- Essien UR, Holmes DN, Jackson LR 2nd, et al. Association of race/ethnicity with oral anticoagulant use in patients with atrial fibrillation: findings from the Outcomes Registry for Better Informed Treatment of Atrial Fibrillation II. *JAMA Cardiol* 2018;3:1174–1182.