Informatics in Medicine Unlocked 25 (2021) 100688

Contents lists available at ScienceDirect

Informatics in Medicine Unlocked

journal homepage: www.elsevier.com/locate/imu

Using machine learning to predict anticoagulation control in atrial fibrillation: A UK Clinical Practice Research Datalink study

Jason Gordon [a,*], Max Norman [a], Michael Hurst [a,c], Thomas Mason [a,b], Carissa Dickerson [a], Belinda Sandler [c], Kevin G. Pollock [c], Usman Farooqui [c], Lara Groves [a], Carmen Tsang [a], David Clifton [d], Ameet Bakhai [e], Nathan R. Hill [c]

[a] Health Economics and Outcomes Research Ltd, Cardiff, UK
[b] School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
[c] Bristol Myers Squibb Pharmaceutical Ltd, Uxbridge, UK
[d] Department of Engineering Science, University of Oxford, Oxford, UK
[e] Cardiology Department, Royal Free London NHS Foundation Trust, London, UK

ARTICLE INFO

Keywords:
Machine learning
Unsupervised learning
Atrial fibrillation
Warfarin
Anticoagulation control
International normalised ratio

ABSTRACT

Objective: To investigate the predictive performance of machine learning (ML) algorithms for estimating anticoagulation control in patients with atrial fibrillation (AF) who are treated with warfarin.

Methods: This was a retrospective cohort study of adult patients (≥18 years) between 2007 and 2016 using linked primary and secondary care data (Clinical Practice Research Datalink GOLD and Hospital Episode Statistics). Various ML techniques were explored to predict suboptimal anticoagulation control, defined as time in therapeutic range (TTR) < 70% based on International Normalised Ratio (INR) 2.0–3.0. Baseline (linear and non-linear support vector machines; random forests; stochastic gradient boosting [XGBoost]; neural networks [NN]) and time-varying data (6-week intervals up to 30 weeks (long-short term memory [LSTM] NN)) were applied. Patient records depicting unique lines of warfarin therapy (LOT) were separated into training (70%) and holdout sets (30%) for model training and testing, respectively.

Results: 35,479 patients were eligible for inclusion, of whom 24,684 and 10,795 were assigned to the training (32,683 unique LOTs) and holdout sets (14,218 unique LOTs). Across all models, depression (diagnosis and/or prescription of antidepressant medication) was a significant driver in predicting anticoagulation control. At baseline, XGBoost was the best-performing model (area under the curve [AUC]: 0.624) due to its ability to identify non-linear associations such as age and weight (greater probability of suboptimal control: <65 and >80 years and <70 kg, respectively). Addition of time-varying data to the LSTM NN improved predictive performance, plateauing at AUC of 0.830 at 30 weeks.

Conclusion: ML algorithms displayed clinically useful ability to predict patients who are at greater risk of suboptimal control. The addition of time-varying data to the algorithm, especially prior INR measurements, improved predictive performance. These algorithms provide improved predictive tools for identifying patients who may benefit from more frequent INR monitoring or switching to alternative therapies.

1. Introduction

Atrial fibrillation (AF) – the most common sustained arrhythmia [1] – is characterised by an irregular cardiac rhythm, which increases the risk of blood pooling in the atria. AF can be further categorised as valvular or non-valvular AF (NVAF) based on the presence or absence respectively of moderate to severe mitral valve disease (usually rheumatic) or a prosthetic heart valve replacement [2]. Such blood pooling increases the risk of blood clots. Consequently, patients with AF have a five-fold increase in the risk of thromboembolic stroke [3] and are more likely to experience more severe stroke than patients without AF, resulting in increased risk of stroke-related morbidity and mortality [4,5]. Therefore, it is recommended that the majority of patients with AF are prescribed oral anticoagulants (OACs) to reduce both the risk of

* Corresponding author. Health Economics and Outcomes Research Ltd., Rhymney House, Unit A Copse Walk, Cardiff Gate Business Park, Cardiff, CF23 8RB, UK.
E-mail address: jason.gordon@heor.co.uk (J. Gordon).

https://doi.org/10.1016/j.imu.2021.100688
Received 4 March 2021; Received in revised form 26 July 2021; Accepted 30 July 2021
Available online 1 August 2021

blood clot formation and the concomitant risk of ischemic stroke [6].

Warfarin is a vitamin K antagonist (VKA) based anticoagulant that has been used clinically since the 1950s [7], and - despite the availability of newer direct OACs (DOACs) [8–11] – remains widely used in the UK today for the prevention of stroke and systemic embolism in patients with NVAF [12–15]. In many cases this is because the patient is unsuitable for DOAC therapy (e.g. because of severe renal impairment), whereas in others it is due to clinician or patient preference [16,17] or drug cost. For warfarin to be clinically effective, most patients are recommended to remain within a narrow therapeutic range (International Normalised Ratio (INR) between 2.0 and 3.0). Due to various dietary and drug interactions with warfarin [18], patients require regular and frequent INR monitoring, which is resource intensive for health care services and a burden to most patients.

Longer-term control of INR can be quantified using time in therapeutic range (TTR), which is determined by establishing the proportion of time a patient is within range, typically achieved using the Rosendaal method [19] with linear interpolation between known INR measurements. Suboptimal INR control is associated with an increased risk of adverse outcomes; the risk of mortality in patients with a TTR <40% is similar to that in patients with AF not receiving anticoagulation therapy [20]. Therefore, methods to accurately predict – at the time of clinician decision – which patients are likely to have suboptimal warfarin control, and thus may be better suited to DOAC therapy or more frequent INR monitoring if on warfarin, would be beneficial to both clinicians and patients.

Conventional statistical methods, such as regression modelling, have been used to identify patient factors associated with high risk of suboptimal anticoagulation control, however, factors such as high unpredictable inter-individual variability in TTR affect the performance of these methods [21,22]. Artificial intelligence (AI) methods such as machine learning (ML) may be more accurate than conventional regression-based models, by being able to identify complex non-linear associations between variables both at a single and at multiple timepoints [23], and are currently being evaluated across multiple disease areas to improve the diagnosis and management of many conditions. Machine learning techniques have been utilised to develop algorithms to improve the screening and detection of AF. These techniques involve either the screening of routinely collected data contained within electronic medical records to identify patients at highest risk of undiagnosed AF who should undergo further screening [23,24], or the application of ML algorithms to electrocardiogram (ECG) traces to detect waveform changes that are associated with AF [25–28]. Machine learning techniques have also been developed and evaluated to optimise warfarin dosing regimens to reduce the risk of adverse drug events associated with the therapy [29–34]. In the studies that evaluated both conventional statistical methods and ML techniques, there was little or no additional benefit of the ML techniques in the optimisation of warfarin dosing [32,33]. However, none of these studies have specifically targeted patients with AF and none included INR control as the primary outcome.

One tool used clinically to predict the risk of poor INR control in patients with AF is the SAMe-TT_2R_2 score [35]. The tool was developed using linear regression to identify clinical variables associated with poor INR control and is used to identify patients less likely to achieve optimal INR control on warfarin who may be better suited to DOAC therapy [6]. However, a meta-analysis of 16 studies evaluating the clinical usefulness of the SAMe-TT_2R_2 score concluded that, whilst the score was able to predict suboptimal INR control (low TTR), its use in individual patients was too limited to be clinically useful [36].

A recently published study utilised ML techniques to develop a clinical outcome prediction model based on serial INR measurements in patients with AF [37]. INR measurements within the first 30 days of treatment were used to predict major bleed, stroke or systemic embolism, and all-cause mortality up to 12 months thereafter. The ML model outperformed TTR in the prediction of adverse clinical outcomes, and whilst there is value in the prediction of adverse clinical events in patients with AF receiving warfarin therapy, there may also be value in methods to identify patients at high risk of suboptimal INR control (who may or may not go on to experience an adverse event).

We are unaware of any tools or algorithms – developed using ML techniques – to predict risk of suboptimal INR control in patients with AF receiving warfarin therapy. Therefore, we sought to evaluate the relative predictive performances of different ML algorithms for estimating anticoagulation control in patients with NVAF treated with warfarin in the UK.

Contributions of this study to knowledge include:

- An evaluation of the performance of different ML methods based on baseline and/or time-varying data in predicting the risk of suboptimal INR control in patients with NVAF and treated with warfarin
- Determination of the key risk drivers at baseline and over time of predicting suboptimal INR control

The remaining sections of the paper are organised as follows: section 2 outlines the study methodology; section 3 presents the results; and section 4 discusses the results and the implications, and limitations of the study.

2. Methods

2.1. Study design

This was a retrospective cohort study of patients with NVAF who were initiated on warfarin between the 1st January 2007 and 31st December 2016 (study period) using routinely collected electronic health records from primary and linked secondary care. Data were gathered from the Clinical Practice Research Datalink (CPRD) GOLD database, which contains anonymised medical records from over 11.3 million patients from GP practices across the UK [38]. CPRD provided primary care data linked with secondary care data from the Hospital Episode Statistics (HES) database [39]; linkage is currently around 60% of GP practices in England participating in CPRD [38]. The study data set also contained linked mortality data from the Office for National Statistics death registrations, and deprivation data at both GP practice level (Index of Multiple Deprivation) and patient level (Townsend Deprivation Index, for England only). CPRD is one of the largest databases of longitudinal medical records derived from primary care in the world [40] and is being used increasingly by European researchers as the demographic characteristics of the UK primary care population are comparable with many European populations [41,42] and thus findings are considered highly generalisable. The study protocol (18_187R) was approved by the Independent Scientific Advisory Committee for the Medicines and Healthcare Products Regulatory Agency on 22 August 2018.

2.2. Patients

Adult patients (≥18 years) were included if they had a Read (primary care) or International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) (secondary care) diagnosis code for AF (or atrial flutter [AF/F]) during the study period, and a subsequent British National Formulary (BNF) code for warfarin treatment after AF diagnosis. Patients were excluded if they met at least one of the following criteria: Read or ICD-10 code for AF/F prior to the study period; Read or ICD-10 code for pulmonary embolism, deep-vein thrombosis, mitral valve disease, or valve surgery prior to AF diagnosis (i.e. the latter two indications excluded patients with valvular AF who are ineligible for DOAC prescribing); BNF code for warfarin >3 months prior to AF diagnosis; BNF code for DOAC treatment prior to warfarin initiation; less than two INR measurements within a six month period and occurring ≥6 weeks after warfarin initiation; registered with

a GP practice participating in CPRD for <12 months prior to AF diagnosis; or, most recent CPRD up-to-standard (UTS) date <12 months prior to AF diagnosis.

2.3. Data management

Data were extracted from the CPRD GOLD database in 2019. Baseline demographics, co-morbidities, body mass index (BMI), alcohol, smoking, prescriptions, primary care interactions, documented CHA_2DS_2-VASc score and related data were extracted. CHA_2DS_2-VASc scores were recalculated using available patient data. In the case of repeated scores, the most recent score was used. Units were standardised prior to data cleaning and data points deemed to be implausible, invalid, or clinically infeasible were removed according to pre-defined data ranges, as listed inSupplementary Table S1. In addition, non-explicit values as recorded in patient notes (e.g. <; >) were excluded. Where possible, imputation of missing values was undertaken, including BMI (if height and weight were known), pulse pressure (if systolic and diastolic pressure were known), total cholesterol (TC) (if high density lipoprotein (HDL) and or low density lipoprotein (LDL) were known), cholesterol ratios (if TC and HDL or LDL were known), and alcohol intake categories (if weekly units were known). The index date for each line of therapy (LOT) was the date of warfarin initiation and each LOT ran until the earliest of 1) 31st December 2016; 2) death; 3) GP practice no longer contributing to CPRD; 4) patient transferring out of GP practice; 5) warfarin discontinuation (≥90-day gap in supply from end of prior prescription end date). If a patient had a subsequent recorded warfarin prescription after a discontinuation criterion was met, this was considered a new line of warfarin therapy. Eligible patients were split into two data sets; the training set (70%) designed to train the model and the holdout (test) set (30%) to evaluate the predictive performance of each model. All splits were performed at patient-level rather than by LOT, so all LOTs were assigned to the same set. 10-fold cross validation was repeated three times to define model hyper parameters, model tuning and feature selection. All analyses were undertaken in R version 3.4 or later.

2.4. Variable definitions

Anticoagulation control was evaluated using individual-level TTR – defined as the proportion of days, post-stabilisation phase, that a patient's INR was calculated to be within 2.0 and 3.0 over each line of warfarin therapy. INR measurements within the first 42-days, considered the stabilisation phase were excluded. TTR was calculated using linear interpolation (based on Rosendaal et al. [19]), and a TTR threshold of <70% was defined as suboptimal INR control [22]. In addition to TTR, anticoagulation control was further categorised based on time spent over the therapeutic range (TOR; i.e. INR >3.0) and time spent under the therapeutic range (TUR; i.e. INR <2.0) [22]. A high TOR places patients at greater risk of bleeding-related events, whilst a significant TUR places patients at greater risk of clotting-related events. A TOR or TUR threshold of >15% was defined as suboptimal INR control, selected as remaining areas either side of the optimal range of >70% TTR [22]. TUR and TOR are not mutually exclusive, and patients with highly variable INR measurements can have both a high TOR and TUR.

2.5. Models to predict anticoagulation control

To develop the predictive models, known risk predictors of anticoagulation control were identified from both the literature [22,43] and clinical expert opinion to create a maximum feature set. These included a range of demographic variables, clinical measures, health behaviours, comorbidities, and medication use, as listed inSupplementary Table S2. Predictive models were developed using both conventional statistical (logistic regression) and ML methods to identify relationships between patient demographic and clinical characteristics at baseline (defined as the date of warfarin initiation) and suboptimal INR control. A time-varying ML model was also developed to explore the effect of both baseline and time-dependent data on relationships between patient characteristics and suboptimal INR control at 6-week intervals following the index date.

For all predictive models (except the neural networks (NN)), a binary measure of INR control (i.e. suboptimal control versus optimal control) was used as the primary outcome. In the NN models, a continuous measure of INR control (i.e. TTR) was used, with a TTR of <70% defined as suboptimal control. Model performance was evaluated using the area under the receiver operating characteristic (AUROC) curve as well as common ML performance values (including sensitivity, specificity, positive predictive value [PPV] and negative predictive value [NPV]) and confusion matrices at 25%, 50% and 75% sensitivities for each model.

2.5.1. Baseline logistic regression model

Logistic regression was applied to baseline data using a binary measure of INR control, with suboptimal INR control set as the positive class. The logistic regression model was applied using the glm package in R. A list of features included in the logistic regression model is detailed inSupplementary Table S2.

2.5.2. Baseline machine learning (ML) models

The following ML techniques were evaluated: Random Forest; XGBoost; linear and non-linear support vector machines (SVM); and NNs. These models were chosen because they cover a breadth of different ML techniques, allowing for both linear and non-linear associations to be assessed.

During model development, a reduced feature set was constructed for each model. This process removed any non-predictive features, along with any features that were highly collinear with other features in the set. The individual feature sets were then consolidated into a single, common feature set. All features were ranked according to either correlation with (logistic regression, linear SVM) or mutual information (non-linear SVM, NN MLP, Random Forest, XGBoost) to the target variable, i.e., suboptimal INR control. The ranked list of features was stepped through sequentially with each feature being added to the model's feature set if the current score improved upon the best score up to that point. The score was defined as the mean AUROC curve found by cross-validation minus two standard errors minus an additional small value, ε. The feature set was then assessed. Any features that were retained but deemed not to be clinically plausible were removed. Any features that were not retained but for which a strong clinical argument could be made for inclusion were evaluated on a case-by-case basis. A list of features included in each baseline model is detailed inSupplementary Table S2. Accumulated local effects (ALE) plots were used to quantitatively describe the influence of features in the model. ALE plots were used as this visual method of model interpretation is unbiased compared with other methods.

2.5.3. Time-varying machine learning (ML) models

A long short-term memory (LSTM) recurrent NN (RNN) was developed to estimate INR control both at baseline and over the first six months of follow-up after warfarin initiation and INR control. The LSTM RNN was chosen because of its ability to examine associations over time [44].

The LSTM RNN was trained on a time-series dataset structured into intervals of six weeks from week 0 (the date of warfarin initiation) to week 30 for all variables with the exception of INR measurements. Time-varying INR measurements were included in the LSTM RNN from week 6 to week 30; INR data from the first 6 weeks after warfarin initiation were not considered, to allow for stabilisation of INR levels. Stabilisation performance varies by clinical choices on starting dose regimes, frequency of initial INR testing and dose changes between clinicians and centers. In addition to estimating TTR, the LSTM time-varying model simultaneously estimated TOR and TUR as three concurrent output

nodes to stabilise the predictions across the three outcomes. As the LSTM was developed over the early stages of patient follow-up, several additional features derived from previously recorded INR measurements post-stabilisation phase were included. Each patient's prior TOR and TUR were included along with the deviation of the patient's INR readings from 2.5, i.e., from the midpoint of the recommended therapeutic range, and the average difference between consecutive INR readings. Additionally, the LSTM model was forced to retain the patients' smoking status and alcohol consumption as weekly units. The final feature set used in the LSTM is detailed in theSupplementary Table S2.

3. Results

3.1. Patient characteristics

35,479 patients were eligible for the study. Analyses were undertaken on 24,684 unique patients (training set) and evaluated on 10,795 unique patients (holdout set). Within the training set, the 24,684 patients contributed 32,683 unique LOTs; 78% (n = 19,279) of patients contributed one LOT, 15% contributed two LOTs, and the remaining 7% contributing three or more LOTs during the 10-year study period. 13,843 (42.4%) of LOTs were classified as suboptimal control and 18,840 (57.6%) were classified as optimal control. Treatment duration was shorter in those LOTs defined as suboptimal control compared to optimal control (620.2 [confidence interval [CI]: 610.5–629.9] vs. 906.1 [CI: 895.5–916.7] days). Across all patients, mean follow-up duration was 778.2 days (standard deviation [SD]: 692.2 days), with a mean of 614.2 days (SD: 583.7) in patients with suboptimal INR control and 903.9 days (SD: 740.9) in patients with optimal INR control.

Demographic and clinical characteristics stratified by INR control are summarised in Table 1. There were no clinically meaningful differences in patient age, BMI, or ethnicity on INR control, however there was a significant gender disparity, with a greater proportion of men exhibiting suboptimal control (43.9% [male] vs 41.4% [female]; $p < 0.001$) based on TTR. Patients with suboptimal control consumed more alcohol 12.0 vs 10.2 units per week ($p < 0.001$), had more frequent contact with healthcare services (10.7 GP visits and 0.42 hospitalisations per year vs 9.46 GP visits and 0.31 hospitalisations per year; $p < 0.001$) and less favourable distributions of both Townsend deprivation index and CHA_2DS_2-VASc scores than patients with optimal INR control ($p < 0.001$).

Patients in the suboptimal INR control group had a higher prevalence of comorbidities such as diabetes, depression, anxiety, anaemia, coronary heart disease, heart failure, pulmonary disease, and renal disease, amongst others (all $p < 0.001$ vs optimal INR control). Patients with suboptimal INR control were also more likely to have received prescriptions for antidepressants, antibiotics, asthma medications, digoxin, diuretics and proton pump inhibitors in the three months prior to warfarin initiation (all $p < 0.001$ vs optimal INR control).

3.2. Baseline logistic regression model

All features included in the logistic regression model demonstrated statistically significant relationships with INR control, with depression being the strongest predictor of suboptimal control. Patients who were either diagnosed with depression or prescribed antidepressants in the three months prior to the index date were ~1.3 times (OR: 1.324; 95% CI: 1.248–1.405) more likely to exhibit suboptimal INR control compared to patients without either of these. Another significant driver of suboptimal INR control was red blood cell count, with a single unit increase associated with a reduction (OR: 0.762; 95% CI: 0.731, 0.795) in the probability of suboptimal INR control (Supplementary Figure S2).

The confusion matrices at 25%, 50%, and 75% sensitivity for the logistic regression model are summarised in Table 2. The AUROC for this model was 0.606. At 75% sensitivity (i.e. the threshold to correctly identify 75% of suboptimal INR control cases), 4,626 (32.5%) LOTs were correctly labelled as suboptimal control (true positives [TP]) while 5,040 were incorrectly labelled as such (false positives [FP]), resulting in positive and negative predictive values (PPV and NPV) of 47.9% and 66.1% respectively.

3.3. Baseline machine learning models

Of the ML models evaluated, the best performing model was XGBoost with an AUROC of 0.624, followed closely by the Random Forest model with an AUROC of 0.621, as shown in Table 2 and Fig. 1. The AUROC for the NN (MLP), SVM (r.b.f.), and SVM (linear) models were 0.618, 0.617, and 0.605 respectively (Supplementary Table S2). In the XGBoost model, at 75% sensitivity, 4,628 (32.6%) LOTs were correctly labelled as suboptimal control (TP) while 4,751 (33.4%) were incorrectly labelled as such (FP), resulting in a PPV and NPV of 49.3% and 68.1% respectively.

Of the binary variables, depression – defined as a diagnosis present at any time in the five years prior to index date or prescription for antidepressant medication in the three months prior to the index date – was the single most important driver in the XGBoost model (ALE = 0.0537), with pulmonary disease (ALE = 0.0295) and asthma (ALE = 0.232) being the next two most important variables (Supplementary Figure S1). Of the continuous variables, alcohol consumption and annual number of hospital admissions had positive linear associations with the probability of suboptimal INR control, and red blood cell count and albumin concentration had negative linear associations (Supplementary Figure S2).

As shown in Fig. 2, the XGBoost model was also able to identify non-linear associations between covariates such as body weight and age and suboptimal INR control. Body weight <70 kg was associated with greater probability of suboptimal control. Patients weighing 50 kg were 6.5% points more likely to exhibit suboptimal INR control compared with patients weighing 70 kg. From 70 to 103 kg the relationship between body weight and INR control was relatively flat. The relationship between patient age and suboptimal INR control was 'U shaped', with patients aged <65 years or >80 years exhibiting a greater probability of suboptimal INR control than patients aged 65–80 years.

3.4. Time-varying machine learning models

The LSTM RNN model incorporated additional time-varying features not included in the baseline ML models, namely: time-dependent smoking status; annual length of hospital stay; CHA_2DS_2-VASc scores; and prior INR measurements (including TOR and TUR in addition to TTR) in the post-stabilisation phase. The AUROC curve and performance statistics for the time-varying LSTM NN model stratified by weeks from index (i.e. weeks from warfarin initiation) are displayed in Table 3. At week 0 (i.e. at baseline), the performance (AUC 0.616) was comparable to other baseline models. As time post-index increased and more information was available, there was a significant increase in model performance, culminating in an AUC of 0.830 at week 30, as demonstrated in Fig. 3.

The confusion matrices at 25%, 50%, and 75% sensitivity levels for the LSTM RNN at Week 30 are summarised in Table 2. At 75% sensitivity, 4,628 (32.6%) LOTs were correctly labelled as suboptimal control (TP) while 2,024 (14.2%) were incorrectly labelled as such (FP), resulting in positive and negative predictive values (PPV and NPV) of 69.6% and 79.6% respectively.

Three different modules were included in the final LSTM model: 1) baseline information, 2) time-varying information excluding INR data and 3) time-varying INR data. The individual contributions of each module to the final model are shown in Table 3. Whilst the performance of the time-varying model without INR data (i.e. only including: smoking status; annual length of hospital stay; and CHA_2DS_2-VASc scores) only improved slightly over time, the inclusion of INR data in the time-varying INR module improved performance to a greater extent. Additional presentation of these results in the form of case studies is

Table 1
Demographic and clinical characteristics of the 24,684 patients included in the training set.

	Suboptimal control* (n = 13,843)		Optimal control* (n = 18,840)		p-value†	
	Mean (SD)	N (%)	Mean (SD)	N (%)	Welch's *t*-test	Chi-square test
Patient demographics (latest value within 5 years prior to LOT initiation)						
Age (years)	74.9 (0.1)	–	74.6 (0.1)	–	0.005	–
Female	–	7,795 (56.1%)	–	11,067 (58.7%)	–	<0.001
Male	–	6,048 (43.9%)	–	7,773 (41.3%)	–	<0.001
Weight (kg)	82.6 (0.2)	–	84.1 (0.2)	–	<0.001	–
Height (m)	1.69 (0.0)	–	1.70 (0.0)	–	<0.001	–
BMI (kg/m2)	28.9 (0.1)	–	29.2 (0.1)	–	<0.001	–
RBC (count)	4.5 (0.5)		4.6 (0.5)		<0.001	
Smoking status					–	<0.001
Non-smoker	–	7,461 (53.9%)	–	10,694 (56.8%)		
Current smoker	–	360 (2.6%)	–	427 (2.3%)		
Ex-smoker	–	6,022 (43.5%)	–	7,719 (41.0%)		
Ethnicity					–	0.009
Black	–	17 (0.1%)	–	28 (0.1%)		
Asian	–	13 (0.1%)	–	11 (0.1%)		
Other	–	255 (1.8%)	–	384 (2.0%)		
South Asian	–	48 (0.3%)	–	31 (0.2%)		
White	–	5,765 (41.6%)	–	7,810 (41.5%)		
Not recorded	–	7,745 (55.9%)	–	10,576 (56.1%)		
Townsend deprivation index					–	<0.001
1 (least deprived)	–	1,980 (14.3%)	–	2,883 (15.3%)		
2	–	2,063 (14.9%)	–	2,996 (15.9%)		
3	–	1,758 (12.7%)	–	2,317 (12.3%)		
4	–	1,329 (9.6%)	–	1,696 (9.0%)		
5 (most deprived)	–	748 (5.4%)	–	810 (4.3%)		
Not recorded‡	–	5,966 (43.1%)	–	8,158 (43.3%)		
Index of multiple deprivation					–	0.279
1 (least deprived)	–	1,680 (12.1%)	–	2,315 (12.3%)		
2	–	1,780 (12.9%)	–	2,431 (12.9%)		
3	–	2,388 (17.3%)	–	3,150 (16.7%)		
4	–	2,050 (14.8%)	–	2,657 (14.1%)		
5 (most deprived)	–	1,809 (13.1%)	–	2,537 (13.5%)		
Not recorded‡	–	4,136 (29.9%)	–	5,750 (30.5%)		
Time (AF diagnosis to warfarin initiation; days)	892.3 (9.8)	–	843.5 (8.3)	–	<0.001	–
CHA2DS2-VASc score^					–	<0.001
0 (lowest risk)	–	584 (4.2%)	–	686 (3.6%)		
1	–	1,170 (8.5%)	–	1,718 (9.1%)		
2	–	2,190 (15.8%)	–	3,446 (18.3%)		
3	–	3,048 (22.0%)	–	4,666 (24.8%)		
4	–	3,235 (23.4%)	–	4,451 (23.6%)		
5	–	2,315 (16.7%)	–	2,704 (14.4%)		
6	–	1001 (7.2%)	–	961 (5.1%)		
7	–	267 (1.9%)	–	187 (1.0%)		
8	–	33 (0.2%)	–	21 (0.1%)		
9 (highest risk)	–	0 (0.0%)	–	0 (0.0%)		
Health behaviour (latest value within 5 years prior to LOT initiation)						
Alcohol (weekly units)	12.0 (0.2)	–	10.2 (0.2)	–	<0.001	–
Alcohol consumption					–	<0.001
Very heavy	–	127 (0.9%)	–	78 (0.4%)		
Heavy	–	186 (1.3%)	–	152 (0.8%)		
Moderate	–	690 (5.0%)	–	761 (4.0%)		
Light	–	5,472 (39.5%)	–	7,883 (41.8%)		
None	–	1,294 (9.3%)	–	1,725 (9.2%)		
Ex	–	1,386 (10.0%)	–	1,799 (9.5%)		
Not recorded	–	4,688 (33.9%)	–	6,442 (34.2%)		
Comorbidity history (any time 5 years prior to LOT initiation)						
Anaemia	–	1,620 (11.7%)	–	1,488 (7.9%)	–	<0.001
Anxiety	–	581 (4.2%)	–	622 (3.3%)	–	<0.001
Bleeding (any)	–	1,537 (11.1%)	–	1,790 (9.5%)	–	<0.001
Bleeding (gastrointestinal)	–	1,481 (10.7%)	–	1,733 (9.2%)	–	<0.001
Bleeding (intracerebral)	–	42 (0.3%)	–	57 (0.3%)	–	1.000
Bleeding (intracranial)	–	55 (0.4%)	–	57 (0.3%)	–	0.365
Cancer	–	1,869 (13.5%)	–	2,261 (12.0%)	–	<0.001
Use of pain medication	–	3,571 (25.8%)	–	3,749 (19.9%)	–	<0.001
Coronary artery disease	–	1,301 (9.4%)	–	1,375 (7.3%)	–	<0.001
Coronary heart disease	–	3,655 (26.4%)	–	4,277 (22.7%)	–	<0.001
Dementia	–	360 (2.6%)	–	320 (1.7%)	–	<0.001
Depression	–	1,952 (14.1%)	–	1,790 (9.5%)	–	<0.001
Deep vein thrombosis	–	221 (1.6%)	–	170 (0.9%)	–	<0.001
Heart failure	–	3,959 (28.6%)	–	4,578 (24.3%)	–	<0.001
Hyperlipidaemia	–	1,800 (13.0%)	–	2,242 (11.9%)	–	0.004
Hypertension	–	8,430 (60.9%)	–	11,134 (59.1%)	–	0.001

(*continued on next page*)

Table 1 (*continued*)

	Suboptimal control* (n = 13,843)		Optimal control* (n = 18,840)		p-value†	
	Mean (SD)	N (%)	Mean (SD)	N (%)	Welch's *t*-test	Chi-square test
Myocardial infarction	–	914 (6.6%)	–	980 (5.2%)	–	<0.001
Osteoarthritis	–	2,949 (21.3%)	–	3,655 (19.4%)	–	<0.001
Pulmonary embolism	–	235 (1.7%)	–	188 (1.0%)	–	<0.001
Pulmonary disease	–	3,613 (26.1%)	–	3,787 (20.1%)	–	<0.001
Renal disease	–	3,807 (27.5%)	–	4,352 (23.1%)	–	<0.001
Stroke	–	1,544 (11.2%)	–	1,945 (10.3%)	–	0.017
TIA	–	1,093 (7.9%)	–	1,556 (8.3%)	–	0.242
Diabetes (T1DM or T2DM)	–	2,118 (15.3%)	–	2,374 (12.6%)	–	<0.001
Tachycardia	–	927 (6.7%)	–	1,074 (5.7%)	–	<0.001
Concomitant medication usage (prescribed within the 3 months prior to LOT initiation)						
ACE inhibitors	–	221 (1.6%)	–	245 (1.3%)	–	0.036
Antibiotics	–	3,405 (24.6%)	–	3,711 (19.7%)	–	<0.001
Antidepressants	–	1,523 (11.0%)	–	1,319 (7.0%)	–	<0.001
Aspirin	–	5,136 (37.1%)	–	7,253 (38.5%)	–	0.009
Asthma medication	–	2,658 (19.2%)	–	2,807 (14.9%)	–	<0.001
Beta blockers	–	2,035 (14.7%)	–	3,090 (16.4%)	–	<0.001
Calcium channel blockers	–	4,333 (31.3%)	–	6,161 (32.7%)	–	0.012
Dementia medication	–	69 (0.5%)	–	57 (0.3%)	–	0.009
Digoxin	–	3,488 (25.2%)	–	3,825 (20.3%)	–	<0.001
Diuretics	–	5,399 (39.0%)	–	5,859 (31.1%)	–	<0.001
Hypertension medication	–	138 (1.0%)	–	170 (0.9%)	–	0.109
Lipid lowering drugs	–	7,475 (54.0%)	–	10,362 (55.0%)	–	0.095
NSAIDs	–	775 (5.6%)	–	980 (5.2%)	–	0.182
Platelet aggregate inhibitors	–	1,163 (8.4%)	–	1,507 (8.0%)	–	0.192
Proton pump inhibitors	–	4,900 (35.4%)	–	5,916 (31.4%)	–	<0.001
Medical contact (within the 5 years prior to LOT initiation)						
GP visits per year (normal)	10.7 (0.1)	–	9.5 (0.1)	–	<0.001	–
GP visits per year (acute)	0.2 (0.0)	–	0.1 (0.0)	–	<0.001	–
Hospital admissions per year	0.4 (0.0)	–	0.3 (0.0)	–	<0.001	–

AF: atrial fibrillation; BMI: body mass index; GP: general practice; LOT: line of therapy; NSAIDs: non-steroidal anti-inflammatory drugs; RBC: red blood cells;TIA: transient ischemic attack.

* Suboptimal/Optimal control defined as time in therapeutic range (TTR) <70/>70% respectively.

†Used for statistical significance testing between suboptimal and optimal control groups.

ˆCHA_2DS_2-VASc scores were recalculated using available patient data. In the case of repeat scores, the most recent score was used.

Table 2
Confusion matrices for the baseline logistic regression model, the best performing machine learning model, and the time-varying machine learning model at weeks 0 and 30 at 25%, 50% and 75% sensitivities.

	AUROC	Sensitivity	Specificity	TP	FP	TN	FN	PPV	NPV
Conventional model									
Logistic Regression (n = 14,218)	0.606	25.0%	86.2%	1,541	1,110	6,938	4,629	58.1%	60.0%
		50.0%	65.5%	3,085	2,776	5,272	3,085	52.6%	63.1%
		75.0%	37.4%	4,626	5,040	3,008	1,544	47.9%	66.1%
Best performing baseline machine learning model									
XGBoost (n = 14,218)	0.624	25.0%	87.7%	1,543	990	7,058	4,627	60.9%	60.4%
		50.0%	67.1%	3,085	2,647	5,401	3,085	53.8%	63.6%
		75.0%	41.0%	4,628	4,751	3,297	1,542	49.3%	68.1%
Time-varying machine learning model at week 0									
LSTM RNN (n = 14,218)	0.616	25.0%	86.6%	1,543	1,075	6,974	4,627	58.9%	60.1%
		50.0%	66.4%	3,085	2,706	5,342	3,085	53.3%	63.4%
		75.0%	39.4%	4,627	4,880	3,168	1,543	48.7%	67.2%
Time-varying machine learning model at week 30									
LSTM RNN (n = 14,218)	0.830	25.0%	98.1%	1,543	152	7,896	4,627	91.1%	63.1%
		50.0%	91.8%	3,085	661	7,387	3,085	82.4%	70.5%
		75.0%	74.8%	4,628	2,024	6,024	1,542	69.6%	79.6%

AUROC: area under the receiver operating characteristics curve; FN: false negative; FP: false positive; LSTM RNN: long short-term memory recurrent neural network; NPV: negative predictive value; PPV: positive predictive value; TN: true negative; TP: true positive; XGBoost: stochastic gradient boosting.

shown inSupplementary Figure S3.

4. Discussion

Suboptimal INR control was present in 42.4% of LOTs in the training set. Patients receiving a LOT defined as suboptimal control were more likely to be male, were more likely to report moderate to heavy alcohol consumption and were more likely to be socioeconomically deprived. Interactions between alcohol and warfarin are well-established [45,46], and – although the effect was more modest in our study – excessive alcohol consumption has been linked to a 3-fold increase in the risk of poor INR control [47]. Furthermore, across all models at baseline, the largest driver of suboptimal INR control was depression-based markers. Patients diagnosed with depression in the five years prior to index or prescribed anti-depressant medication in the three months prior to warfarin initiation were approximately 30% (based on the logistic regression model) more likely to exhibit suboptimal INR control than patients without either of these markers. These findings support those

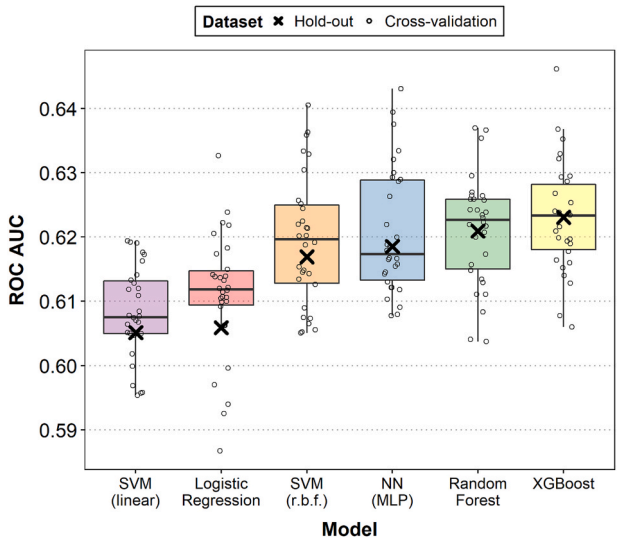

Fig. 1. Box plot outlining the performance of each simulation involved in cross-validation (; training set) and the final performance (× ; holdout set).
AUC: area under the curve; MLP: multi-layer perceptron; NN: neural network; RBF: radial basis function; ROC: receiver operating characteristics; SVM: support vector machine; XGBoost: stochastic gradient boosting.

from other studies that have linked anxiety and depression to reduced INR stability and therefore decreased TTR [48,49], placing these patients at increased risk of either thromboembolic or bleeding events.

The majority of ML models evaluated in this study performed numerically better than the conventional statistical (logistic regression) approach for predicting suboptimal INR control in patients with AF. Unlike conventional statistical methods, ML models can understand complex interrelationships between covariates, with some able to identify non-linear associations. Indeed, the best performing ML model using baseline data – the XGBoost – was able to identify non-linear associations between covariates such as age and weight, potentially a reason for its improved performance compared with linear methods. Overall, whilst these results indicate an improvement in the ability of most ML algorithms to predict patients at higher risk of suboptimal anticoagulation control over conventional methods using baseline data, the incremental predictive ability of the best ML model remains moderate and reflects the complexity and challenges of predicting patients at highest risk of suboptimal INR control with currently documented primary care data.

In contrast to the ML models based on baseline data alone, time-varying ML models allowed for the incorporation of additional information during follow-up after warfarin initiation. In this study, the introduction of time-varying data, in particular prior INR measurements in the post-stabilisation phase, to the LSTM RNN significantly improved the performance of this ML algorithm to predict suboptimal anticoagulation control. Whilst the main driver of this improvement in performance over time was the availability of prior INR measurements, it was not just the INR measurements themselves that contributed to the improved power to predict risk of future suboptimal INR control, but also the relationships between INR measurements and other covariates. As evidenced in the case studies in Figure S3, the ML algorithm identified patients at higher risk of future suboptimal control despite prior optimal control and vice versa. In these highlighted cases, despite patients having a history of optimal control, they were identified as having higher risk of future poor INR control (and conversely patients with history of suboptimal control were identified as having lower risk of future poor control), which was subsequently confirmed by the data. In these cases, clinician assessment of prior INR measurements alone may have led to the assumption that these patients were at lower and higher risk of suboptimal control respectively, despite the opposite finding.

The time-varying ML model (LSTM RNN) demonstrated excellent predictive ability (AUROC >0.8) from week 24 onwards. Optimal therapeutic efficacy of warfarin is realised when INR is maintained

Table 3
AUROC performance for final LSTM RNN time-varying model and performance of disaggregated components – stratified by weeks from index.

Weeks from index	0	6	12	18	24	30
Final model	0.616	0.638	0.703	0.763	0.804	0.830
Disaggregated modules						
Baseline information only	0.620					
Time-varying (without INR)	0.605	0.607	0.613	0.619	0.627	0.632
Time-varying INR	0.500	0.576	0.674	0.746	0.793	0.822

INR: international normalised ratio

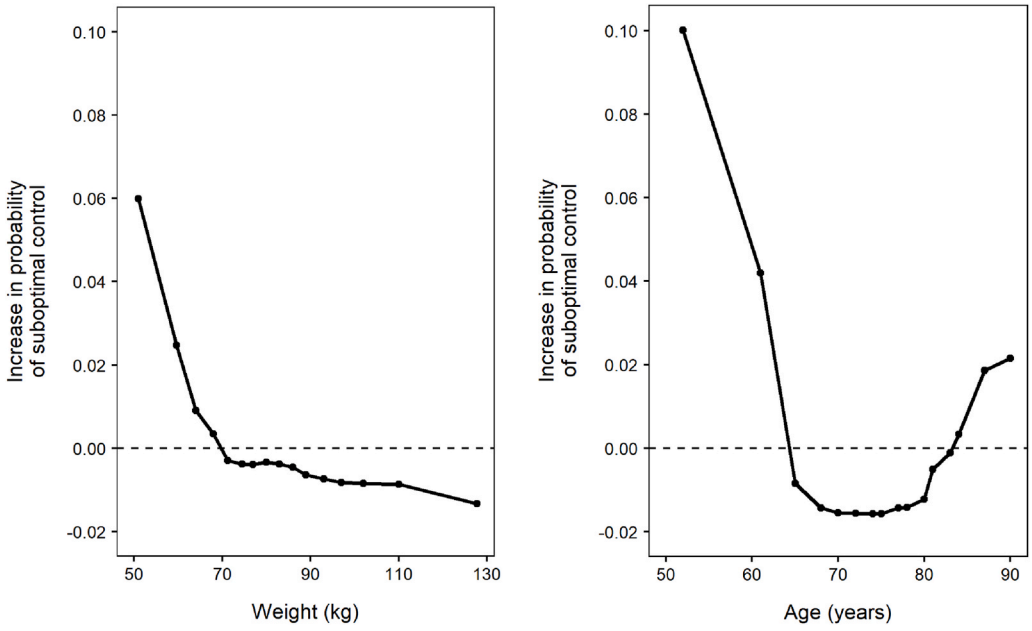

Fig. 2. (Left) change in probability of suboptimal control (based on accumulated local effects, ALE) for weight for the stochastic gradient boosting (XGBoost) baseline model and (Right) change in probability of suboptimal control against age for the XGBoost baseline model.
(Left) At baseline: <70 kg–6,794 patients; ≥70 kg and <110 kg–23,387 patients; ≥110 kg–2,395 patients (Right) At baseline: <60 years–2,335; ≥60 years and <85 years–26,354; ≥85 years–3,887 patients.

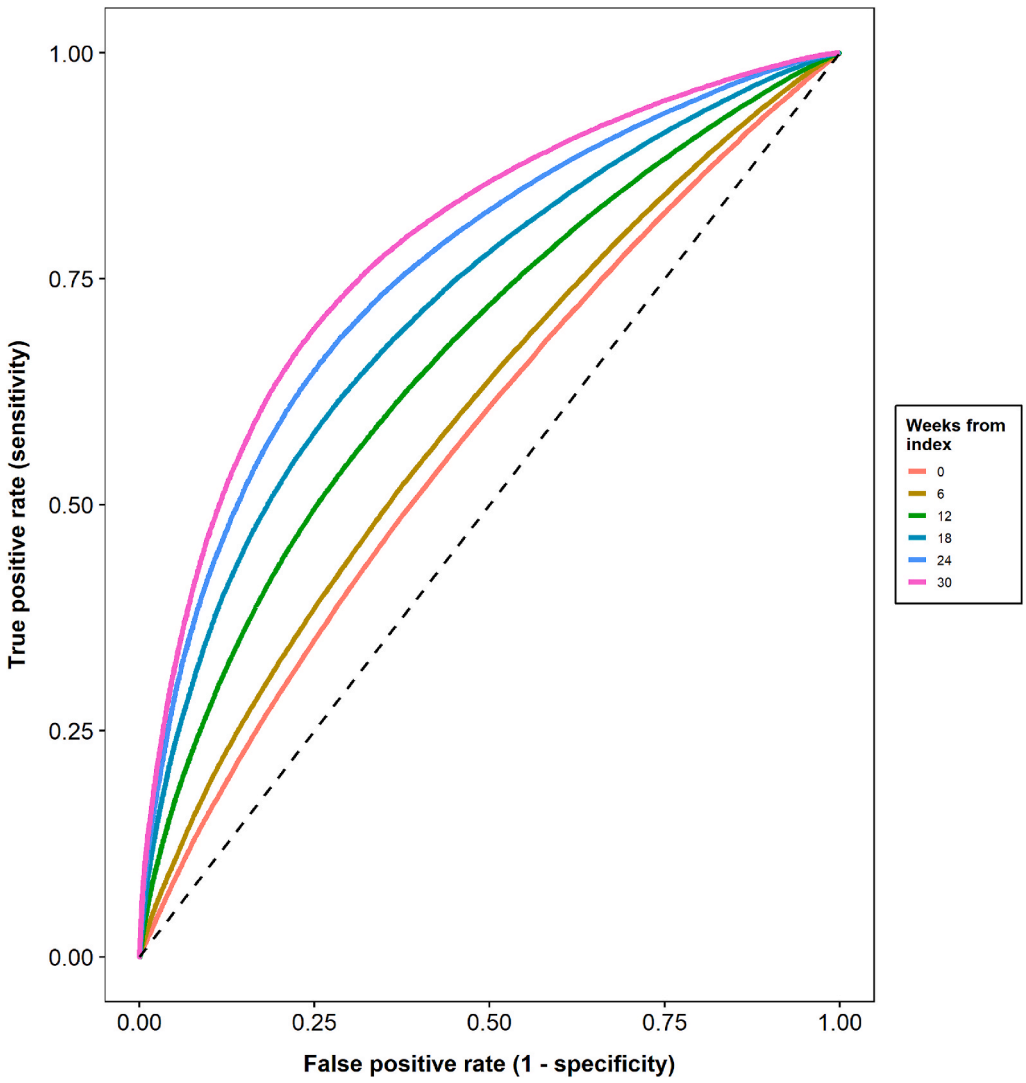

Fig. 3. Receiver operating characteristics (ROC) curve for the time-varying long short-term memory recurrent neural network (LSTM RNN) model.

within 2.0–3.0. However, to ensure INR is maintained within this narrow range, patients must regularly attend warfarin clinics. Time-varying ML methods could be utilised as an additional tool in warfarin clinics to routinely assess the likelihood of individual patients experiencing suboptimal INR control in the future, based on their current time-updated profile. Patients identified as being at higher risk of suboptimal control could be offered more frequent INR monitoring or switched to an alternative therapy prior to the onset of suboptimal control. This is especially pertinent during the COVID-19 pandemic, where many patients are being switched from warfarin to DOACs to improve the safety of anticoagulation monitoring in line with national guidelines.

ML techniques have previously been studied as a means of optimising warfarin dosage [29,30] and for the prediction of clinical outcomes [37]. However, the only published study that we are aware of to report on the development of a method to predict TTR or the quality of oral anticoagulation (the SAMe-TT_2R_2 score) in patients with AF did not employ ML methods [35]. Studies have reported variable ability of the SAMe-TT_2R_2 score to predict patients at risk of poor INR control, and as a result, a recent meta-analysis of 16 studies concluded there was limited clinical utility of the SAMe-TT_2R_2 score to predict poor INR control at individual patient level [36]. Whilst we have not directly compared the SAMe-TT_2R_2 score with the ML algorithms applied in this study, our results indicate that there is benefit in using ML methods over conventional statistical approaches to predict patients at higher risk of suboptimal INR control, especially with the inclusion of time-varying data. However, whether ML methods can be developed to accurately predict which patients are at higher risk of suboptimal INR control at baseline prior to warfarin initiation requires further investigation in the real-world clinical setting.

A major strength of this study is that data were obtained from large databases of real-world primary care and secondary care activity. Whilst the inclusion of >30,000 patients increases the robustness of associations between patient-related variables and suboptimal INR control and its generalisability to other populations, the study is not without limitations. First, as this is a retrospective database study, analyses and conclusions are reliant on the accuracy and completeness of original data entry. CPRD take steps to ensure this by only providing data of research quality, but nevertheless errors are likely to still exist. Second, only patient records from approximately 60% of general practices that contribute to CPRD GOLD are linked with HES, meaning there was incomplete linkage with secondary care records across the entire dataset. Third, selection bias may also have occurred because we only included patients who had sufficient follow-up data (i.e. at least two INR measurements in the first six months after warfarin initiation). Lastly, dietary (for example: vitamin K based foods) data, emerging parameters in anticoagulation response such as genetic variants and performance of the anticoagulation services responsible for each patient, and HAS-BLED score (to quantify major bleeding risk) – which were first described in

2010 but not used commonly in clinical practice until after 2014 [50,51] – were unavailable for input into our models.

In conclusion, at baseline, ML models demonstrated better predictive ability of INR control compared to conventional methods. The inclusion of time-varying data (notably prior INR measurements) significantly improved the performance of the ML methods in predicting suboptimal INR control, however, it is not just the INR measurements themselves but the complex interplay between INR measurements and other clinical characteristics that predict risk. Such time-varying ML algorithms may be useful as an additional tool in warfarin clinics to routinely assess the likelihood of a patient experiencing suboptimal INR control in the future.

Funding

This study was sponsored by Bristol Myers Squibb Pharmaceutical Ltd & Pfizer Inc.. The funding agreement ensured the authors' independence in designing the study, interpreting the data, and preparing the manuscript for publication.

Author contributors

G, UF, BS, NH, LG, DC, and AB conceptualised and designed the study. MN, MH, and TM were responsible for data analysis. All authors contributed to interpretation of the results, preparation and review of the manuscript, and approval of the final manuscript for publication.

Consent

Data collection was entirely retrospective and patient data was anonymised by the Clinical Practice Research Datalink (CPRD) prior to delivery to the authors. There was no direct patient contact and written informed consent was not required.

Data availability

Data cannot be shared publicly because they are the property of Clinical Practice Research Datalink (CPRD). Data are available from CPRD under a license agreement.

Declaration of competing interest

BS, KP, UF, and NH are employed by Bristol Myers Squibb (BMS) UK Ltd. JG, MN, CD, LG, and CT are current employees of Health Economics and Outcomes Research (HEOR) Ltd. who received funding from BMS UK Ltd. for this study. MH and TM were employees of HEOR Ltd. at the time of the study; MH is now employed by BMS UK Ltd. DC undertakes consultancy work with Biobeats Ltd. and Sensyne Health plc. AB does not hold any stock but is involved in research sponsored by and is a member of advisory panels and speakers' bureau for Daiichi Sankyo, Pfizer, BMS, Bayer and Boehringer Ingelheim.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.imu.2021.100688.

References

[1] Atrial Fibrillation Association Anticoagulation Europe (UK). The AF report - atrial fibrillation: preventing A stroke crisis2012 26 January 2018. Available from: http://www.preventaf-strokecrisis.org/.

[2] Camm AJ, Lip GY, De Caterina R, Savelieva I, Atar D, Hohnloser SH, et al. Focused update of the ESC Guidelines for the management of atrial fibrillation: an update of the 2010 ESC Guidelines for the management of atrial fibrillation. Developed with the special contribution of the European Heart Rhythm Association. Eur Heart J 2012;33(21):2719–47. 2012.

[3] Wolf P, Abbott R, Kannel W. Atrial fibrillation as an independent risk factor for stroke: the Framingham Study. Stroke 1991;22(8):983–8.

[4] Lamassa M, Di Carlo A, Pracucci G, Basile AM, Trefoloni G, Vanni P, et al. Characteristics, outcome, and care of stroke associated with atrial fibrillation in Europe: data from a multicenter multinational hospital-based registry (The European Community Stroke Project). Stroke 2001;32(2):392–8.

[5] Marini C, De Santis F, Sacco S, Russo T, Olivieri L, Totaro R, et al. Contribution of atrial fibrillation to incidence and outcome of ischemic stroke: results from a population-based study. Stroke 2005;36(6):1115–9.

[6] Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomstrom-Lundqvist C, et al. ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): the Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J 2020;42(5):373–498. 2021.

[7] Wardrop D, Keeling D. The story of the discovery of heparin and warfarin. Br J Haematol 2008;141(6):757–63.

[8] National Institute for Health and Care Excellence. Technology appraisal guidance [TA249]: Dabigatran etexilate for the prevention of stroke and systemic embolism in atrial fibrillation [Available from, https://www.nice.org.uk/guidance/ta249.

[9] National Institute for Health and Care Excellence. Technology appraisal guidance [TA256]: Rivaroxaban for the prevention of stroke and systemic embolism in people with atrial fibrillation [Available from: https://www.nice.org.uk/guidance/ta256.

[10] National Institute for Health and Care Excellence. Technology appraisal guidance [TA275]: Apixaban for preventing stroke and systemic embolism in people with nonvalvular atrial fibrillation [Available from: https://www.nice.org.uk/guidance/ta275.

[11] National Institute for Health and Care Excellence. Technology appraisal guidance [TA355]: edoxaban for preventing stroke and systemic embolism in people with non-valvular atrial fibrillation [Available from: https://www.nice.org.uk/guidance/ta355.

[12] De Caterina R, Kelly P, Monteiro P, Deharo JC, de Asmundis C, Lopez-de-Sa E, et al. Characteristics of patients initiated on edoxaban in Europe: baseline data from edoxaban treatment in routine clinical practice for patients with atrial fibrillation (AF) in Europe (ETNA-AF-Europe). BMC Cardiovasc Disord 2019;19(1):165.

[13] Ho KH, van Hove M, Leng G. Trends in anticoagulant prescribing: a review of local policies in English primary care. BMC Health Serv Res 2020;20(1):279.

[14] Kongnakorn T, Lanitis T, Annemans L, Thijs V, Goethals M, Marbaix S, et al. Stroke and systemic embolism prevention in patients with atrial fibrillation in Belgium: comparative cost effectiveness of new oral anticoagulants and warfarin. Clin Drug Invest 2015;35(2):109–19.

[15] Le Heuzey JY, Ammentorp B, Darius H, De Caterina R, Schilling RJ, Schmitt J, et al. Differences among western European countries in anticoagulation management of atrial fibrillation. Data from the PREFER IN AF registry. Thromb Haemostasis 2014;111(5):833–41.

[16] Alonso-Coello P, Montori VM, Diaz MG, Devereaux PJ, Mas G, Diez AI, et al. Values and preferences for oral antithrombotic therapy in patients with atrial fibrillation: physician and patient perspectives. Health Expect 2015;18(6):2318–27.

[17] Ikeda T, Yasaka M, Kida M, Imura M. A survey of reasons for continuing warfarin therapy in the era of direct oral anticoagulants in Japanese patients with atrial fibrillation: the SELECT study. Patient Prefer Adherence 2018;12:135–43.

[18] Holbrook AM, Pereira JA, Labiris R, McDonald H, Douketis JD, Crowther M, et al. Systematic overview of warfarin and its drug and food interactions. Arch Intern Med 2005;165(10):1095–106.

[19] Rosendaal FR, Cannegieter SC, van der Meer FJ, Briet E. A method to determine the optimal intensity of oral anticoagulant therapy. Thromb Haemostasis 1993;69(3):236–9.

[20] Morgan CL, McEwan P, Tukiendorf A, Robinson PA, Clemens A, Plumb JM. Warfarin treatment in patients with atrial fibrillation: observing outcomes associated with varying levels of INR control. Thromb Res 2009;124(1):37–41.

[21] Francis B, Lane S, Pirmohamed M, Jorgensen A. A review of a priori regression models for warfarin maintenance dose prediction. PloS One 2014;9(12):e114896.

[22] Macedo AF, Bell J, McCarron C, Conroy R, Richardson J, Scowcroft A, et al. Determinants of oral anticoagulation control in new warfarin patients: analysis using data from Clinical Practice Research Datalink. Thromb Res 2015;136(2):250–60.

[23] Hill NR, Ayoubkhani D, McEwan P, Sugrue DM, Farooqui U, Lister S, et al. Predicting atrial fibrillation in primary care using machine learning. PloS One 2019;14(11):e0224582.

[24] Tiwari P, Colborn KL, Smith DE, Xing F, Ghosh D, Rosenberg MA. Assessment of a machine learning model applied to harmonized electronic health record data for the prediction of incident atrial fibrillation. JAMA Netw Open 2020;3(1):e1919396.

[25] Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction. Lancet 2019;394(10201):861–7.

[26] Lown M, Brown M, Brown C, Yue AM, Shah BN, Corbett SJ, et al. Machine learning detection of Atrial Fibrillation using wearable technology. PloS One 2020;15(1):e0227401.

[27] Oster J, Hopewell JC, Ziberna K, Wijesurendra R, Camm CF, Casadei B, et al. Identification of patients with atrial fibrillation: a big data exploratory analysis of the UK Biobank. Physiol Meas 2020;41(2):025001.

[28] Poh MZ, Poh YC, Chan PH, Wong CK, Pun L, Leung WW, et al. Diagnostic assessment of a deep learning system for detecting atrial fibrillation in pulse waveforms. Heart 2018;104(23):1921–8.
[29] Ma Z, Wang P, Gao Z, Wang R, Khalighi K. Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. PloS One 2018;13(10):e0205872.
[30] Sharabiani A, Bress A, Douzali E, Darabi H. Revisiting warfarin dosing using machine learning techniques. Comput Math Methods Med 2015;2015:560108.
[31] Hu YH, Wu F, Lo CL, Tai CT. Predicting warfarin dosage from clinical data: a supervised learning approach. Artif Intell Med 2012;56(1):27–34.
[32] Li X, Liu R, Luo ZY, Yan H, Huang WH, Yin JY, et al. Comparison of the predictive abilities of pharmacogenetics-based warfarin dosing algorithms using seven mathematical models in Chinese patients. Pharmacogenomics 2015;16(6):583–90.
[33] Nguyen VL, Nguyen HD, Cho YS, Kim HS, Han IY, Kim DK, et al. Comparison of multivariate linear regression and a machine learning algorithm developed for prediction of precision warfarin dosing in a Korean population. J Thromb Haemostasis 2021;19(7):1676–86.
[34] Roche-Lima A, Roman-Santiago A, Feliu-Maldonado R, Rodriguez-Maldonado J, Nieves-Rodriguez BG, Carrasquillo-Carrion K, et al. Machine learning algorithm for predicting warfarin dose in caribbean hispanics using pharmacogenetic data. Front Pharmacol 2019;10:1550.
[35] Apostolakis S, Sullivan RM, Olshansky B, Lip GYH. Factors affecting quality of anticoagulation control among patients with atrial fibrillation on warfarin: the SAMe-TT(2)R(2) score. Chest 2013;144(5):1555–63.
[36] van Miert JHA, Bos S, Veeger N, Meijer K. Clinical usefulness of the SAMe-TT2R2 score: a systematic review and simulation meta-analysis. PloS One 2018;13(3):e0194208.
[37] Goto S, Goto S, Pieper KS, Bassand JP, Camm AJ, Fitzmaurice DA, et al. New artificial intelligence prediction model using serial prothrombin time international normalized ratio measurements in atrial fibrillation patients on vitamin K antagonists: GARFIELD-AF. Eur Heart J Cardiovasc Pharmacother 2020;6(5):301–9.
[38] Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T, et al. Data resource profile: clinical practice research Datalink (CPRD). Int J Epidemiol 2015;44(3):827–36.
[39] NHS Digital. Hospital Episode statistics (HES) 2019 [Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics.
[40] Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the general practice research database as an example of a UK primary care data resource. Ther Adv Drug Saf 2012;3(2):89–99.
[41] de Jong RG, Gallagher AM, Herrett E, Masclee AA, Janssen-Heijnen ML, de Vries F. Comparability of the age and sex distribution of the UK Clinical Practice Research Datalink and the total Dutch population. Pharmacoepidemiol Drug Saf 2016;25(12):1460–4.
[42] Ibanez L, Sabate M, Vidal X, Ballarin E, Rottenkolber M, Schmiedl S, et al. Incidence of direct oral anticoagulant use in patients with nonvalvular atrial fibrillation and characteristics of users in 6 European countries (2008-2015): a cross-national drug utilization study. Br J Clin Pharmacol 2019;85(11):2524–39.
[43] Williams BA, Evans MA, Honushefsky AM, Berger PB. Clinical prediction model for time in therapeutic range while on warfarin in newly diagnosed atrial fibrillation. J Am Heart Assoc 2017;6(10).
[44] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[45] Wells PS, Holbrook AM, Crowther NR, Hirsh J. Interactions of warfarin with drugs and food. Ann Intern Med 1994;121(9):676–83.
[46] Weathermon R, Crabb DW. Alcohol and medication interactions. Alcohol Res Health 1999;23(1):40–54.
[47] Bjorck F, Kadhim H, Sjalander A. Predictors for INR-control in a well-managed warfarin treatment setting. J Thromb Thrombolysis 2019;47(2):227–32.
[48] Diug B, Evans S, Lowthian J, Maxwell E, Dooley M, Street A, et al. The unrecognized psychosocial factors contributing to bleeding risk in warfarin therapy. Stroke 2011;42(10):2866–71.
[49] Michal M, Prochaska JH, Ullmann A, Keller K, Gobel S, Coldewey M, et al. Relevance of depression for anticoagulation management in a routine medical care setting: results from the ThrombEVAL study program. J Thromb Haemostasis 2014;12(12):2024–33.
[50] Marcucci M, Lip GY, Nieuwlaat R, Pisters R, Crijns HJ, Iorio A. Stroke and bleeding risk co-distribution in real-world patients with atrial fibrillation: the Euro Heart Survey. Am J Med 2014;127(10):979–986 e2.
[51] Pisters R, Lane DA, Nieuwlaat R, de Vos CB, Crijns HJ, Lip GY. A novel user-friendly score (HAS-BLED) to assess 1-year risk of major bleeding in patients with atrial fibrillation: the Euro Heart Survey. Chest 2010;138(5):1093–100.