



# One Year Risk Prediction Model for Cardiovascular Disease for Adults with Asthma in England

Jessica Baskaran, Vesselin A Novov, Jennifer K Quint , Constantinos Kallis 

School of Public Health, Imperial College London, London, United Kingdom

Correspondence: Constantinos Kallis, School of Public Health, Floor 9, Sir Michael Uren Building, 86 Wood Ln, London, W12 0BZ, United Kingdom, Email [c.kallis@imperial.ac.uk](mailto:c.kallis@imperial.ac.uk)

**Background:** Cardiovascular disease (CVD) continues to be a significant health threat to humans globally, and a significant burden on healthcare systems. Cardiovascular risk prediction utilizing machine learning (ML) models in patients with asthma remains vastly underexplored.

**Methods:** In this cohort study consisting of 641,042 participants, we used routinely collected electronic healthcare record data to explore various ML algorithms including logistic regression, penalized logistic regression, decision trees, random forest and gradient boost to develop a model with high specificity.

**Results:** The penalized logistic regression model was identified to be the best and simplest classification model in terms of discriminatory power (AUC = 0.85). The gradient boost model was found to be the best predictive model in terms of calibration where the predicted and observed probabilities at risk of CVD match or are closely aligned. In all models, the number of previous cardiovascular events was the most influential predictor, followed by age and prescriptions related to cardiovascular medications. The top predictor alone produced a reasonable level of predictive power (AUC = 0.66).

**Conclusion:** We have created a novel prediction model for predicting CVD within a year of asthma diagnosis for patients with asthma at least 50 years old. Using penalized logistic regression, we achieved a high level of accuracy. By implementing this model, it would be possible to screen out patients with low risk of CVD with high specificity and acceptable sensitivity. Penalized logistic regression and gradient boost models have similar accuracy in screening out individuals at low risk of CVD. For this objective, penalized logistic regression may be more suitable than gradient boost models for implementation as it is simpler to use and more transparent. At the probability threshold of 8% (outcome prevalence), both models' effectiveness in reducing unnecessary treatments was by approximately 52%. These ML models performed better compared with traditional statistical-based risk prediction models. The unique contribution of the study is the construction of prediction models for CVD disease within 12 months from asthma diagnosis based on regression and machine learning models and the comparison of their accuracy to identify the best model based on suitable statistical measures such as AUC and calibration. Further prospective studies using different populations and external validation are required to assess and validate the ML risk prediction models.

**Keywords:** asthma, cardiovascular, risk prediction, electronic healthcare records

## Introduction

Cardiovascular diseases (CVD) are among the leading causes of morbidity and mortality globally. In England, approximately 6.4 million people are living with a heart or circulatory disease.<sup>1</sup> And given the benefits of primary prevention, it is essential to implement methods for early detection and prevention of CVD.

Asthma is a common chronic respiratory condition affecting 5.4 million people in the UK and has been identified as a significant risk factor for CVD.<sup>2</sup>

Both CVD and asthma share some common risk factors such as obesity, smoking and physical inactivity, which can lead to the exacerbation of both these events simultaneously. Furthermore, certain medications which are prescribed to treat asthma. Beta agonists and inhaled corticosteroids may potentially impact cardiovascular health.

At present, there are established approaches to risk assessment of CVD. A comprehensive guide has been developed by the American Heart Association and the American College of Cardiology (AHA/ACC).<sup>3</sup> Well-established risk factors include age, hypertension, cholesterol, diabetes, and smoking status.<sup>3</sup> Examples of established approaches include Framingham Risk Score (FRS), Atherosclerotic Cardiovascular Disease (ASCVD) Risk Calculator and QRISK.<sup>4</sup> Whilst these tools assist in identifying individuals at risk for CVD, there still remains a large proportion of individuals who are not identified by these tools and hence preventative treatments are not provided to those at risk. From previous studies, approximately 50% of myocardial infarctions and stroke are expected to occur in individuals who are less likely to be at risk of CVD.<sup>5</sup>

There are a few studies which have explored predicting CVD risk in patients with COPD. ML approaches include logistic regression (LR), support vector machines (SVM), decision trees (DT), random forests (RF), gradient boost (GB), K-Nearest Neighbor (KNN) and neural networks. Some of these methods have been applied in predicting liver and skin diseases.<sup>6,7</sup> There has been an ML study which was conducted in the aim of improving CVD risk prediction using routine clinical data obtained from the Clinical Practice Research Datalink (CPRD) comparing four classes of ML algorithms. In comparison to the AHA/ACC risk prediction algorithm, the four ML algorithms developed had an increased accuracy in identifying individuals who are more likely to develop CVD.<sup>3</sup>

We explored a range of potential predictors available from routinely collected electronic healthcare record data to develop a simple personalized risk prediction model for cardiovascular events within 12 months from asthma diagnosis date.

We have included patients with asthma at least 50 years old as 50 years is a sensible age when cardiovascular disease is more likely to become symptomatic. Intervention would be required to reduce the risk of major life-threatening (MACE) events for patients at high risk.

## Methods

### Data Sources

This study used routinely collected patient data from the Clinical Practice Research Datalink (CPRD) Aurum database. This database is a large, anonymised and longitudinal primary care database of patient electronic healthcare records (EHR) for over 45 million patients captured as of September 2023 from 1,737 primary care practices.<sup>8</sup> For this study, the CPRD Aurum was linked with secondary data including the Hospital Episode Statistics (HES) Accident and Emergency and HES Admitted Patient Care (APC) from English NHS healthcare providers.<sup>9</sup> CPRD Aurum was also linked with data from the Office for National Statistics (ONS) Death Registration, and it is considered the gold standard for mortality data.<sup>9</sup> Furthermore, the CPRD dataset linked with the Index of Multiple Deprivation data.

### Study Design

We defined a cohort of people with a record of GP diagnosis of between the periods of 1<sup>st</sup> January 2010 and 31<sup>st</sup> December 2019. The index date was defined as the initial date of asthma diagnosis recorded in primary care records. The definition of asthma used in this study has been validated in the CPRD GOLD dataset, demonstrating a high positive predictive value (PPV) of >86%.<sup>10,11</sup>

We included patients aged at least 50 years old at the time of asthma diagnosis as they are at a higher risk to develop a CVD event in comparison to younger individuals. All patients selected for study are registered with a GP practice eligible for data linkage to HES and ONS data.

Patients included had to meet data quality standards, be male or female and being registered with a CPRD practice for at least one year. Patients were followed until the earliest date of the following: date of death, CPRD registration end date, CPRD practice last collection date, patient turning 100 years old or the study end date of 31<sup>st</sup> December 2019 or date of last HES and ONS data collection.

### Predictor Variables and Outcome

Predictor variables were selected based on clinical expertise and related scientific literature on asthma and CVD.

We included demographic variables; age (continuous variable), sex and geographical region collated from GP practice data. A number of clinical variables were also related to asthma and CVD events and treatments. Specifically, we included separately the number of asthma exacerbations, the number of SABA (Short-Acting Beta Agonist) and the number of ICS (Inhaled Corticosteroids) canister prescriptions during the last 12 months before index date. In addition, we have included a binary indicator for having at least one GP review for asthma in the last 12 months before index date. In terms of prescriptions related to asthma, we have included a categorical predictor corresponding to GINA steps (1–5) criteria combined with separate categories that corresponded to patients with SABA only prescriptions and to patients with no medications related to asthma during the last 12 months.<sup>12</sup> Prescriptions of medications related to CVD in the last 12 months from asthma index date included indicators for having at least one of the following (separately): positive inotropes, diuretics, anti-arrhythmic, beta-blockers, hypertension and heart failure, anti-anginals, anti-coagulants, anti-platelets and statins. We have also included an indicator for having at least one cardiovascular event in the last 12 months before asthma index date. Co-morbid conditions were also incorporated including prior history of atopy, Gastro-Esophageal Reflux Disease (GORD), Chronic Obstructive Pulmonary Disease (COPD), diabetes, anxiety and depression. Additionally, we included separate indicators for having hypertension and influenza vaccination in last 12 months before asthma index date. Finally, we have added separate indicators for smoking status and vaping status on asthma index date corresponding to current, previous or never smoker and current, previous or never vaper, respectively.

Our outcome variable was defined with a binary indicator for having at least one cardiovascular event recorded within 12 months from their first asthma diagnosis in the study period. Our outcome definition included the following cardiovascular events: acute coronary syndrome, stroke, arrhythmia, and heart failure. The cardiovascular events were recorded and collated from primary care, HES APC and from ONS mortality data. [Supplementary 1](#) contains a summary table of variable descriptions presented above.

## Machine-Learning Algorithms

The study population was split into train (70%), validation (15%) and test (15%) cohorts. Our study compared 4 classes of ML algorithms: logistic regression (LR),<sup>13</sup> decision trees (DT),<sup>14</sup> random forest (RF),<sup>15</sup> and gradient boosting machines (GB)<sup>16</sup> to develop risk prediction models. These algorithms were selected due to the ease of implementation into healthcare systems. If models with equal performance exist, the model that is simpler and more interpretable would be preferable. For the development of each model, the hyperparameters were tuned utilizing the grid search with 5-fold cross validation algorithm to conduct an exhaustive search. Hyperparameters were selected and optimized to achieve a high Area Under ROC Curve (AUC-ROC) for each ML model. The optimal parameters used for each model can be seen in [Supplementary 2](#).

Moreover, to take into account that the number of asthma patients without CVD significantly outnumbered asthma patients with CVD, class weights were implemented.<sup>17</sup> We re-trained the models with weights where higher weights were assigned to asthma patients with CVD (minority class).

The development of the risk prediction models was completed using Python programming language, version 3.11.7 on Visual Studio Code (VSC) facilitating sci-kit learn library packages.

## Feature Importance Analysis

For each of the 5 ML models, we ranked all the features contributing to the prediction of CVD risk to evaluate feature importance in ML algorithms. Feature importance was explored to ensure feature rankings are interpretable and align with clinical expertise.

Utilizing the top 10 features extracted, our study evaluated the performance of each model in comparison to the models with all features.

## Model Evaluation Metrics

Each model was evaluated initially using the AUC-ROC, with a threshold of 0.7. This threshold is an established measure to assess the discriminatory ability between patients at high and low risk for cardiovascular disease.<sup>18,19</sup>

In addition to AUC-ROC, we also estimated other commonly used ML statistical measures such as the confusion matrix and other five-evaluation metrics including accuracy, sensitivity (recall), specificity, precision (also known as positive predictive value) and F1 Score using the test set.<sup>20</sup>

## Calibration

To further assess the performance of models, we produced calibration curves to ensure that the predicted risk levels reflect the actual risk levels, as recommended by Tripod AI guidelines.<sup>21</sup> Calibration curves plot the predicted probabilities against the observed outcomes to determine whether the predicted probabilities obtained from the model are well-calibrated. Calibration curves were produced using Python version 3.11.7 on VSC.

To quantify calibration, slope and Calibration-In-The-Large (CILT, also known as calibration intercept) were estimated using R-studio version 4.3.3.<sup>22</sup>

## Decision Curve Analysis

Decision Curve Analysis (DCA) is also recommended in Tripod AI guidelines. This method evaluates the clinical utility of predictive models.<sup>23</sup> DCA was utilized to obtain the net benefit of a predictive model across a range of threshold probabilities. The threshold was chosen mainly due to the estimate of prevalence. In addition, this threshold would result in a good balance between sensitivity and specificity which is desirable from a clinical point of view. The net benefit was calculated by comparing the true positive and false positive number of patients, weighted by the relative harm of a false positive compared to a false negative. Additionally, the number of net interventions avoided was also examined at a range of thresholds. This number corresponds to the reduction in unnecessary clinical reviews per number of patients when using the model compared to reviewing all or no patients for cardiovascular disease.

DCA was implemented using Python version 3.11.7 on VSC.

## Sensitivity Analysis

To assess the robustness of our CVD risk prediction models, we systematically excluded specific subgroups of patients. Firstly, we excluded all asthma patients who had experienced a previous cardiovascular event 12 months prior from index date. In the second sensitivity analysis, we excluded all asthma patients with a prior history of COPD.

## Ethics

This study complies with the Declaration of Helsinki.

## Results

### Study Population Characteristics

From our cohort of 641,042 participants, there were 50,643 cardiovascular events (8%) within 12 months from index date. About 40% of patients were male among patients without CVD compared to 47% of patients that were male among patients with at least one CVD event. Median age of CVD patients was 74 years (IQR 65–82) compared to median age 61 years (IQR 53–70) for patients with no cardiovascular diagnosis. Further descriptive analysis of characteristics and test statistics are presented in [Table 1](#) and [Supplementary 3](#). In [Table 2](#), descriptive statistics are shown for each sample type (train, validation and test).

### Descriptive Logistic Regression

A logistic regression model was developed estimating the odds ratios (OR) with the corresponding 95% confidence intervals (CI) and p-values to identify the significant predictors of CVD events ([Figure 1](#)). In comparison to males, females had lower odds of having CVD events (OR 0.77, 95% CI [0.75–0.79]). Moreover, co-morbid conditions and medications associated with cardiovascular conditions have also been identified as crucial factors in predicting CVD events. Detailed LR description can be found in [Supplementary 4](#).

**Table 1** Descriptive Statistics and Comparisons by Outcome Status in Training Sample – for Categorical Variables, Numbers and Percentages are Presented with Chi-Square Tests Performed. For Continuous Variables, Medians and Interquartile Ranges (IQR) are Shown Alongside Mann–Whitney Non-Parametric Tests. P-values are Shown for All Variables

Outcome	Without Outcome (n= 413,266)		With Outcome (n= 35,464)		Statistical Test	P-value
	n	%	n	%		
<b>At least one Cardiovascular event within 12 months Predictors</b>						
<b>Gender (n, %)</b>					543.3	<0.001
Female	246,105	60.0	18,870	53.0		
Male	167,161	40.0	16,594	47.0		
<b>Smoking status (n, %)</b>					392.9	<0.001
Current smoker	98,249	24.0	7,762	22.0		
Ex-smoker	237,911	58.0	22,253	63.0		
Never	77,106	19.0	5,449	15.0		
<b>Vaping status (n, %)</b>					142.0	<0.001
Current vaper	10,134	2.5	515	1.5		
Ex-vaper	468	0.1	34	<0.1		
Never-vaper	402,664	97.0	34,915	98.0		
Age on index date	61	53.0–70.0	74	65.0–82.0	3,742,565,693	<0.001
Number of asthma exacerbations 12 months prior	0	0.0–1.0	0	0.0–1.0	6,110,361,093	<0.001
Number of SABA prescriptions 12 months prior	2	0.0–5.0	3	1.0–7.0	6,445,863,711	<0.001
Number of ICS prescriptions 12 months prior	2	0.0–6.0	3	0.0–8.0	6,576,819,413	<0.001
Prior history of atopy (n, %)	186,830	45.0	14,639	41.0	203.9	<0.001
Prior history of GORD (n, %)	77,701	19.0	8,259	23.0	424.6	<0.001
Prior history of COPD (n, %)	66,675	16.0	10,946	31.0	4954.6	<0.001
Prior history of anxiety (n, %)	68,407	17.0	6,288	18.0	32.7	<0.001
Prior history of depression (n, %)	92,950	22.0	7,929	22.0	0.3	0.600
At least 1 GP review for asthma 12 months prior (n, %)	97,940	24.0	8,232	23.0	4.3	0.039
<b>Geographical region in England (n, %)</b>					178.6	<0.001
East Midlands	10,247	2.5	876	2.5		
East of England	20,379	4.9	1,729	4.9		
London	61,365	15.0	4,870	14.0		
North-East	14,224	3.4	1,302	3.7		
North-West	80,813	20.0	7,735	22.0		
South-East	86,191	21.0	6,741	19.0		
South-West	53,655	13.0	4,589	13.0		
West Midlands	70,483	17.0	6,166	17.0		
Yorkshire and The Humber	15,909	3.8	1,456	4.1		
<b>Index of Multiple Deprivation quintiles (n, %)</b>					356.7	<0.001
1 (least deprived)	88,247	21.0	6,604	19.0		
2	87,157	21.0	7,058	20.0		
3	80,167	19.0	6,726	19.0		
4	79,431	19.0	7,136	20.0		
5 (most deprived)	78,264	19.0	7,940	22.0		
<b>GINA/SABA/No medications (n, %)</b>					1826.9	<0.001
Gina step 1	125,541	30.0	8,625	24.0		
Gina step 2	4,479	1.1	279	0.8		
Gina step 3	37,666	9.1	2,999	8.5		
Gina step 4	76,322	18.0	7,465	21.0		
Gina step 5	35,825	8.7	5,118	14.0		
No medications	54,919	13.0	4,096	12.0		
Saba only	78,514	19.0	6,882	19.0		

(Continued)

**Table 1** (Continued).

Outcome	Without Outcome (n= 413,266)		With Outcome (n= 35,464)		Statistical Test	P-value
	n	%	n	%		
Influenza vaccination in last 12 months (n, %)	255,823	62.0	28,421	80.0	4678.7	<0.001
Hypertension in last 12 months (n, %)	46,007	11.0	6,549	18.0	1699.0	<0.001
Diabetes before asthma index date (n, %)	35,963	8.7	6,100	17.0	2776.8	<0.001
Positive inotropes prescriptions 12 months prior (n, %)	2,275	0.6	3,964	11.0	26,903.0	<0.001
Diuretics prescriptions in last 12 months (n, %)	66,128	16.0	15,027	42.0	15,332.0	<0.001
Anti-arrhythmic prescriptions 12 months prior (n, %)	1,430	0.3	1,763	5.0	9889.7	<0.001
Beta-blockers prescriptions in last 12 months (n, %)	24,858	6.0	9,041	25.0	17,744.0	<0.001
Hypertension and heart failure prescriptions 12 months prior (n, %)	111,683	27.0	18,691	53.0	10,449.0	<0.001
Anti-anginals prescriptions in last 12 months (n, %)	77,393	19.0	14,378	41.0	9554.4	<0.001
Anti-coagulants prescriptions 12 months prior (n, %)	8,145	2.0	7,916	22.0	39,194.0	<0.001
Anti-platelets prescriptions in last 12 months (n, %)	53,265	13.0	14,359	40.0	19,439.0	<0.001
Statins prescriptions in last 12 months (n, %)	105,194	25.0	18,262	51.0	11,105.0	<0.001
Previous cardiovascular event 12 months prior (n, %)	7,892	1.9	11,928	34.0	77,861.0	<0.001

**Table 2** Descriptive Statistics and Comparisons by Sample Type – for Categorical Variables, Count and Percentages are Shown with Chi-Square Tests Performed. For Continuous Variables, Medians and Interquartile Ranges (IQR) are Presented Alongside Kruskal–Wallis Non-Parametric Test. P-values are Given for All Variables

Variable	Training Sample (n= 448,730)		Validation Sample (n= 96,156)		Testing Sample (n= 96,156)	
	n	%	n	%	n	%
<b>Outcome</b>						
<b>At least one cardiovascular event within 12 months</b>	35,464	7.9	7,624	7.9	7,555	7.9
<b>Predictors</b>						
<b>Gender (n, %)</b>						
Female	264,975	59.0	56,497	59.0	56,865	59.0
Male	183,755	41.0	39,659	41.0	39,291	41.0
<b>Smoking status (n, %)</b>						
Current smoker	106,011	24.0	22,471	23.0	22,691	24.0
Ex-smoker	260,164	58.0	55,927	58.0	55,791	58.0
Never	82,555	18.0	17,758	18.0	17,674	18.0
<b>Vaping status (n, %)</b>						
Current vaper	10,649	2.4	2,237	2.3	2,297	2.4
Ex-vaper	502	0.1	103	0.1	111	0.1
Never-vaper	437,579	98.0	93,816	98.0	93,748	97.0
Age on index date	62	53.0–72.0	62	53.0–72.0	62	53.0–72.0
Number of asthma exacerbations 12 months prior	0	0.0–1.0	0	0.0–1.0	0	0.0–1.0
Number of SABA prescriptions 12 months prior	2	1.0–5.0	2	1.0–5.0	2	1.0–5.0
Number of ICS prescriptions 12 months prior	2	0.0–6.0	2	0.0–6.0	2	0.0–6.0
Prior history of atopy (n, %)	201,469	45.0	42,994	45.0	43,128	45.0
Prior history of GORD (n, %)	85,960	19.0	18,112	19.0	18,364	19.0
Prior history of COPD (n, %)	77,621	17.0	16,607	17.0	16,488	17.0
Prior history of anxiety (n, %)	74,695	17.0	15,728	16.0	15,992	17.0
Prior history of depression (n, %)	100,879	22.0	21,666	23.0	21,735	23.0
At least one GP review for asthma (n, %)	106,172	24.0	22,576	23.0	22,780	24.0

(Continued)

Table 2 (Continued).

Variable	Training Sample (n= 448,730)		Validation Sample (n= 96,156)		Testing Sample (n= 96,156)	
	n	%	n	%	n	%
<b>Geographical region in England (n, %)</b>						
East Midlands	11,123	2.5	2,388	2.5	2,376	2.5
East of England	22,108	4.9	4,754	4.9	4,863	5.1
London	66,235	15.0	14,440	15.0	14,267	15.0
North-East	15,526	3.5	3,319	3.5	3,367	3.5
North-West	88,548	20.0	18,783	20.0	18,787	20.0
South-East	92,932	21.0	19,844	21.0	19,922	21.0
South-West	58,244	13.0	12,360	13.0	12,416	13.0
West Midlands	76,649	17.0	16,581	17.0	16,371	17.0
Yorkshire and The Humber	17,365	3.9	3,687	3.8	3,787	3.9
<b>Index of Multiple Deprivation quintiles (n, %)</b>						
1 (least deprived)	94,851	21.0	20,282	21.0	20,322	21.0
2	94,215	21.0	20,218	21.0	20,192	21.0
3	86,893	19.0	18,773	20.0	18,484	19.0
4	86,567	19.0	18,594	19.0	18,713	19.0
5 (most deprived)	86,204	19.0	18,289	19.0	18,445	19.0
<b>GINA/SABA/No medications (n, %)</b>						
No medications	59,015	13.0	12,628	13.0	12,609	13.0
Saba only	85,396	19.0	18,143	19.0	18,324	19.0
Gina step 1	134,166	30.0	28,897	30.0	28,793	30.0
Gina step 2	4,758	1.1	958	1.0	1,042	1.1
Gina step 3	40,665	9.1	8,901	9.3	8,719	9.1
Gina step 4	83,787	19.0	17,911	19.0	17,957	19.0
Gina step 5	40,943	9.1	8,718	9.1	8,712	9.1
Influenza vaccination in last 12 months (n, %)	284,244	63.0	60,879	63.0	61,143	64.0
Hypertension in last 12 months (n, %)	52,556	12.0	11,066	12.0	11,145	12.0
Diabetes before asthma index date (n, %)	42,063	9.4	9,126	9.5	9,010	9.4
Positive inotropes prescriptions 12 months prior (n, %)	6,239	1.4	1,343	1.4	1,312	1.4
Diuretics prescriptions in last 12 months (n, %)	81,155	18.0	17,369	18.0	17,170	18.0
Anti-arrhythmic prescriptions in last 12 months (n, %)	3,193	0.7	716	0.7	650	0.7
Beta-blockers prescriptions in last 12 months (n, %)	33,899	7.6	7,190	7.5	7,129	7.4
Hypertension and heart failure prescriptions (n, %)	130,374	29.0	27,739	29.0	27,610	29.0
Anti-anginals prescriptions in last 12 months (n, %)	91,771	20.0	19,681	20.0	19,446	20.0
Anti-coagulants prescriptions in last 12 months (n, %)	16,061	3.6	3,470	3.6	3,404	3.5
Anti-platelets prescriptions in last 12 months (n, %)	67,624	15.0	14,480	15.0	14,333	15.0
Statins prescriptions in last 12 months (n, %)	123,456	28.0	26,350	27.0	26,259	27.0
Previous cardiovascular event in last 12 months (n, %)	19,820	4.4	4,311	4.5	4,241	4.4
Table 2B – Statistical Tests and P-values calculated						
<b>Variable</b>					<b>Statistical Test</b>	<b>P-value</b>
<b>Outcome</b>						
<b>At least one cardiovascular event within 12 months</b>					0.4	0.800
<b>Predictors</b>						

(Continued)

**Table 2** (Continued).

Variable	Statistical Test	P-value
<b>Gender (n, %)</b> Female Male	3.5	0.200
<b>Smoking status (n, %)</b> Current smoker Ex-smoker Never	2.9	0.600
<b>Vaping status (n, %)</b> Current vaper Ex-vaper Never-vaper	1.3	0.900
Age on index date	0.6	0.700
Number of asthma exacerbations 12 months prior	6.3	0.044
Number of SABA prescriptions 12 months prior	2.8	0.200
Number of ICS prescriptions 12 months prior	3.1	0.200
Prior history of atopy (n, %)	1.1	0.600
Prior history of GORD (n, %)	5.3	0.072
Prior history of COPD (n, %)	1.3	0.500
Prior history of anxiety (n, %)	4.9	0.088
Prior history of depression (n, %)	0.72	0.700
At least one GP review for asthma (n, %)	1.6	0.400
<b>Geographical region in England (n, %)</b> East Midlands East of England London North-East North-West South-East South-West West Midlands Yorkshire and The Humber	13.8	0.600
<b>Index of Multiple Deprivation quintiles (n, %)</b> 1 (least deprived) 2 3 4 5 (most deprived)	5.0	0.800
<b>GINA/SABA/No medications (n, %)</b> No medications Saba only Gina step 1 Gina step 2 Gina step 3 Gina step 4 Gina step 5	10.0	0.600

(Continued)

**Table 2** (Continued).

Variable	Statistical Test	P-value
Influenza vaccination in last 12 months (n, %)	2.2	0.300
Hypertension in last 12 months (n, %)	3.8	0.200
Diabetes before asthma index date (n, %)	1.3	0.500
Positive inotropes prescriptions 12 months prior (n, %)	0.5	0.800
Diuretics prescriptions in last 12 months (n, %)	2.8	0.200
Anti-arrhythmic prescriptions in last 12 months (n, %)	3.2	0.200
Beta-blockers prescriptions in last 12 months (n, %)	2.6	0.300
Hypertension and heart failure prescriptions (n, %)	5.3	0.071
Anti-anginals prescriptions in last 12 months (n, %)	2.7	0.300
Anti-coagulants prescriptions in last 12 months (n, %)	0.7	0.700
Anti-platelets prescriptions in last 12 months (n, %)	1.7	0.400
Statins prescriptions in last 12 months (n, %)	1.9	0.400
Previous cardiovascular event in last 12 months (n, %)	0.9	0.600

## Machine Learning Models

All features listed in [Table 2](#) were included into the ML models and trained on 448,730 participants. All ML models produced an AUC-ROC above 0.7. The imbalanced models with unequal class distributions had higher accuracy, specificity and precision in all ML algorithms. To counteract class imbalance, class weights were utilized, but no significant difference in AUC-ROC was observed. All ML models produced 0.45 for Precision-Recall Area Under the Curve (PR-AUC) scores.

## Top 10 Feature Models

The top 10 risk factor features for cardiovascular risk prediction models are presented in [Figure 2](#). For all models, the most influential predictor was “Previous cardiovascular event in last 12 month”. All models identified similar risk factors in feature selection, but with some discrepancies in the order of the rankings.

The performance of the ML models for the top 10 features ([Supplementary 5](#)) had very similar results to the ML models with all features ([Table 3](#)).

## Stepwise Addition of Features

The top feature “previous cardiovascular event” alone produced a reasonable level of predictive power with an AUC of 0.66 in all models ([Figure 3](#)). A significant jump in AUC-ROC has occurred with the feature “Age” highlighting “Age” as an important feature which greatly enhances the model’s discriminatory ability in DT, RF and GB models. No significant specificity variations were observed in all trained and tested models.

## Calibration

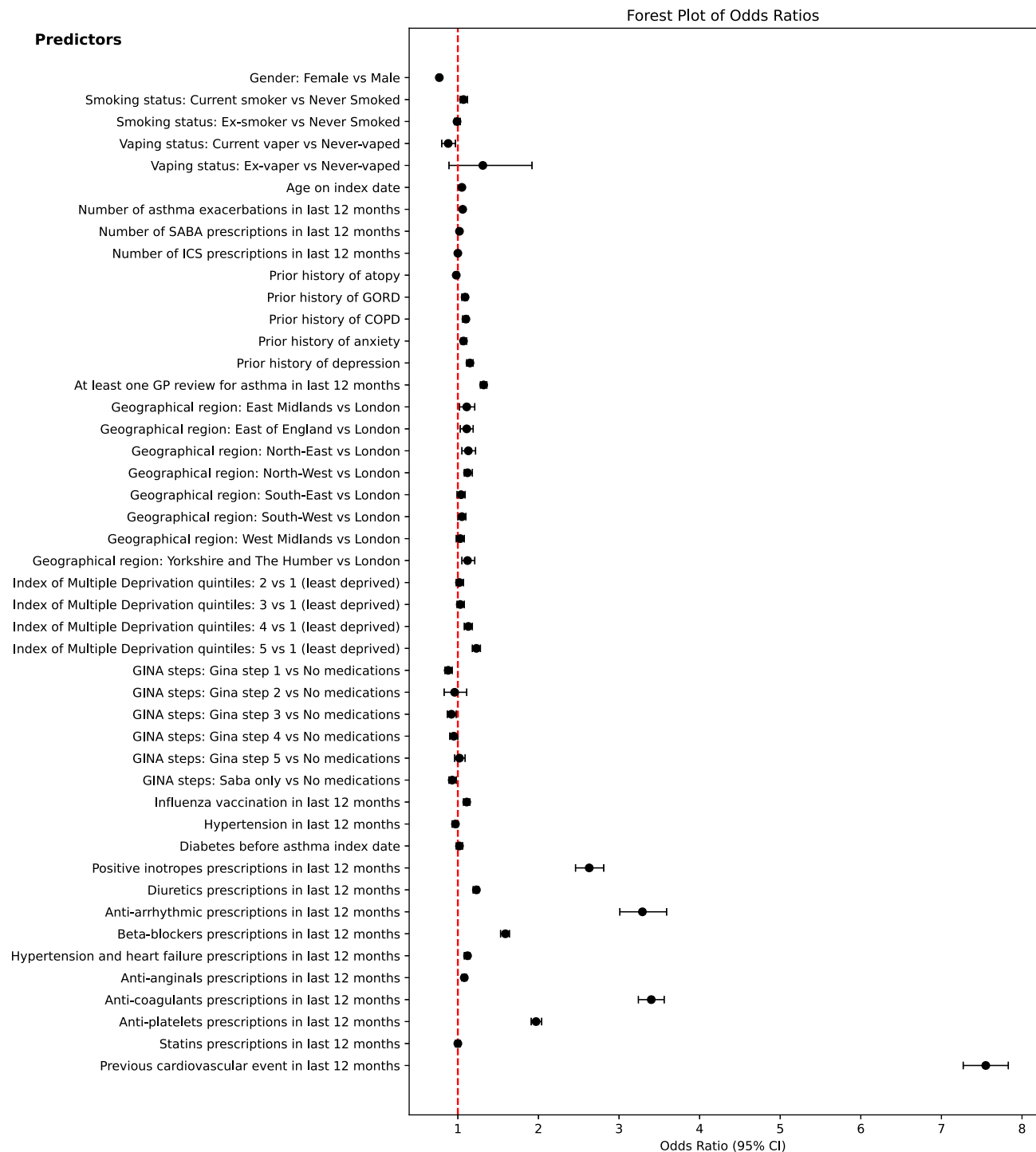
Calibration curves of the full ML models ([Supplementary 6](#)) and reduced top 10 features ([Supplementary 7](#)) had some differences in the predicted risk levels against the actual risk levels. The imbalanced class distributed penalized LR and GB models had slightly better calibration curves in comparison to the other models ([Figure 4](#)).

For the penalized LR utilizing all the features, the CITL was 0.00, and the slope was 1.00, presenting perfect calibration with no systematic bias on unseen (validation and test) data.

Similarly, the GB model including all features illustrated a well-calibrated prediction.

## Decision Curve Analysis

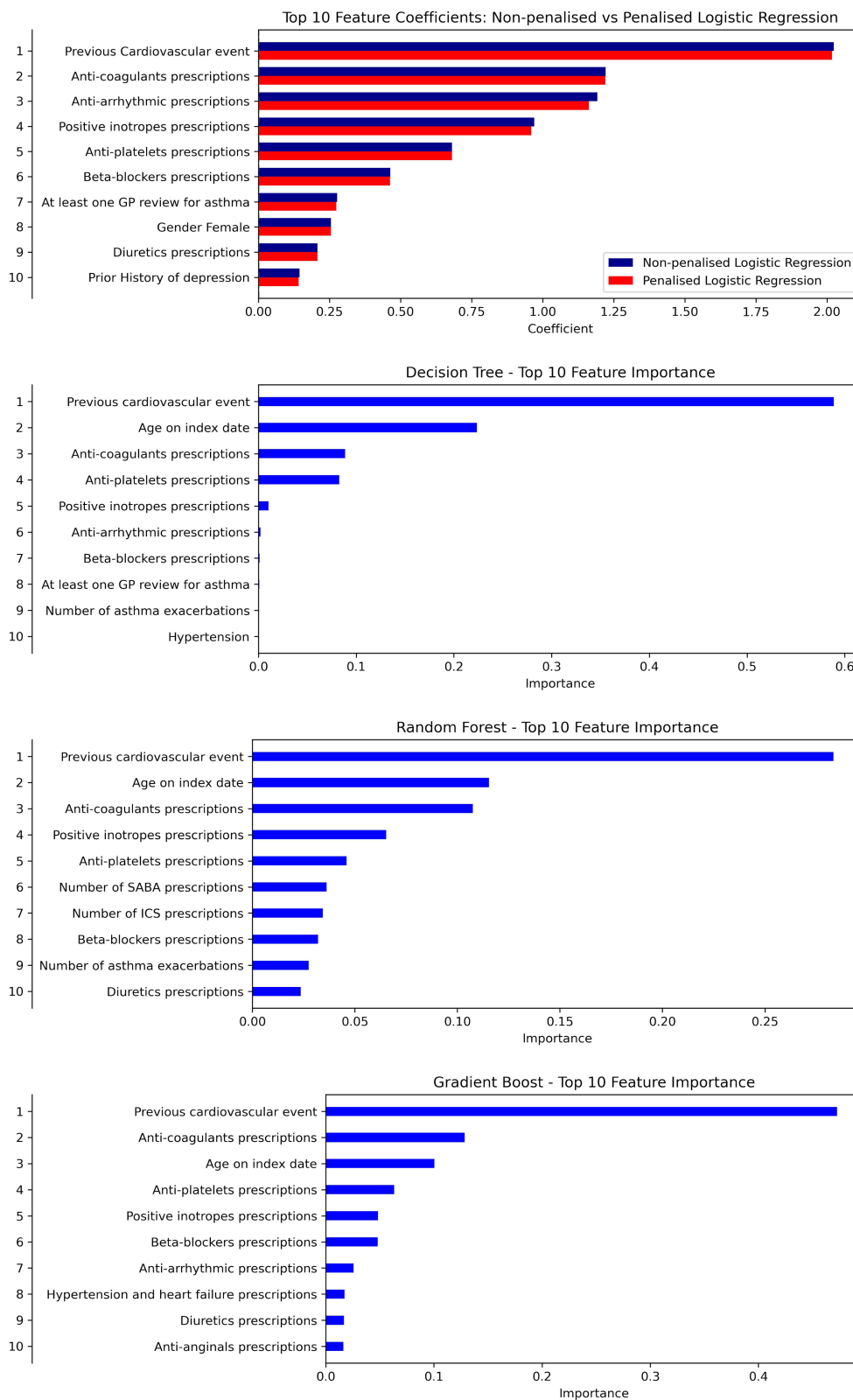
For penalized LR and GB models, clinical utility of the models were assessed through estimation from the net benefit and the net reduction of interventions. About 0.08 (8%) was chosen as a suitable threshold for the model’s predicted



**Figure 1** Forest Plot for the Odds Ratio – Odds Ratio (OR) and 95% confidence intervals (CI) for all predictors are displayed for the outcome variable of at least 1 cardiovascular event within 12 months since index date.

probabilities which was based on outcome prevalence and clinical expertise (Figure 5). In the testing sample, 7,555 asthma patients (7.9%) were identified to have at least 1 CVD event within 12 months since index date.

In the test sample of both penalized LR and GB, the net benefit was 0.04 at the threshold probability of 8%. This means that in comparison to conducting clinical reviews, reviewing on the basis of the predictors is equivalent to identifying 4 CVD events per hundred patients without conducting any unnecessary reviews. Similarly, at probability



**Figure 2** Top 10 Risk Prediction Features for Cardiovascular Disease – The top 10 feature rankings for CVD algorithms have been presented in descending order. Non-penalised Logistic Regression (blue) and Penalised Logistic Regression (red) had the same feature ranking and has been put in the same plot. Logistic Regression use coefficients for feature ranking. Decision Tree, Random Forest and Gradient Boost use importance for feature ranking.

**Table 3** Performance of ML Models for Cardiovascular Risk Prediction Amongst Asthma Patients – Higher AUC Values Result in Better Algorithm Discrimination. Higher Specificity Values Result in Correctly Identifying Individuals Not at Risk for CVD

Models	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
<b>Non-penalized Logistic Regression</b>						
Imbalanced	0.93	0.27	0.99	0.64	0.38	0.85
Weights	0.82	0.73	0.82	0.26	0.38	0.85
Random Under Sampling	0.82	0.72	0.83	0.26	0.38	0.85
<b>Penalized Logistic Regression (Elastic Net)</b>						
Imbalanced	0.93	0.27	0.99	0.64	0.38	0.85
Weights	0.82	0.73	0.82	0.26	0.38	0.85
Random Under Sampling	0.82	0.72	0.83	0.26	0.38	0.85
<b>Decision Tree</b>						
Imbalanced	0.93	0.33	0.98	0.61	0.43	0.84
Weights	0.79	0.75	0.79	0.24	0.36	0.84
Random Under Sampling	0.79	0.75	0.79	0.24	0.36	0.84
<b>Random Forest</b>						
Imbalanced	0.93	0.29	0.99	0.65	0.40	0.86
Weights	0.88	0.60	0.91	0.36	0.45	0.85
Random Under Sampling	0.78	0.77	0.79	0.23	0.36	0.86
<b>Gradient Boost</b>						
Imbalanced	0.93	0.32	0.98	0.63	0.42	0.86
Weights	0.80	0.75	0.8	0.24	0.37	0.86
Random Under Sampling	0.80	0.75	0.8	0.24	0.37	0.86

threshold of 8%, reviewing asthma patients on the basis of the predictors is equivalent to a strategy leading to 51% (penalized LR) and 52% (GB) reduction in the number of clinical reviews without missing any cardiovascular events.

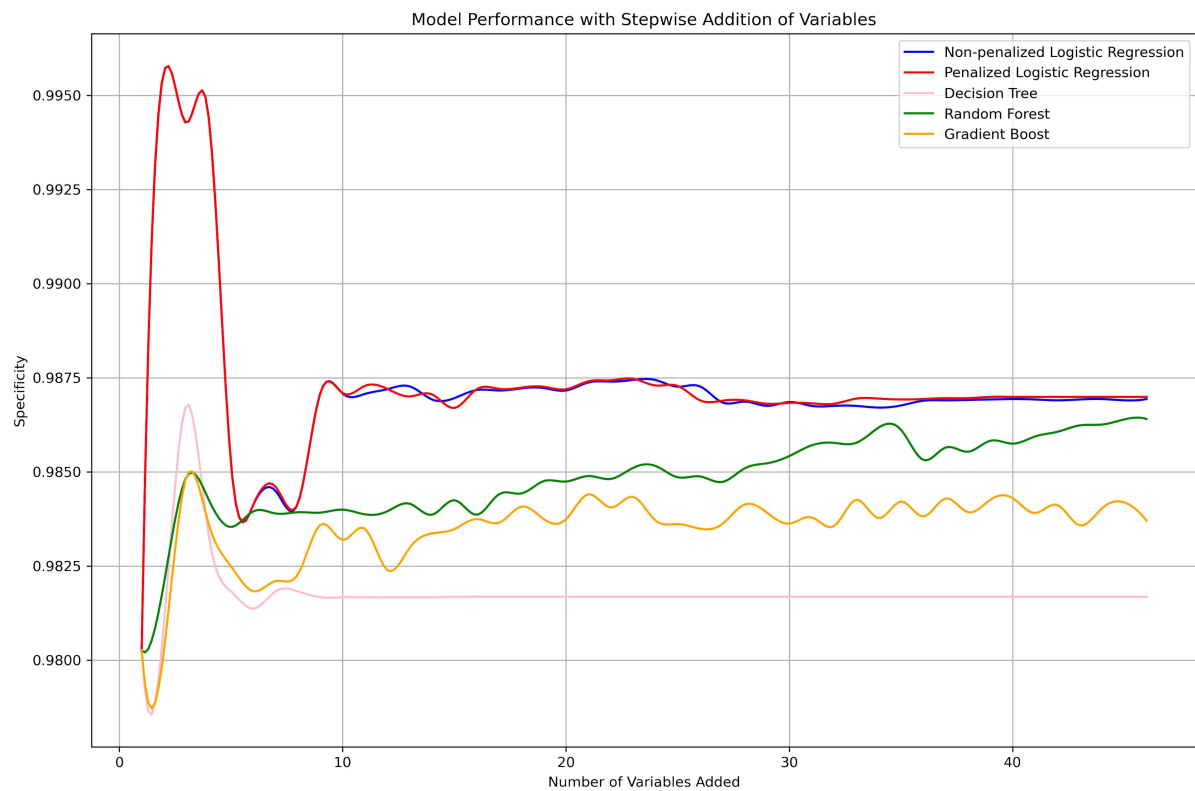
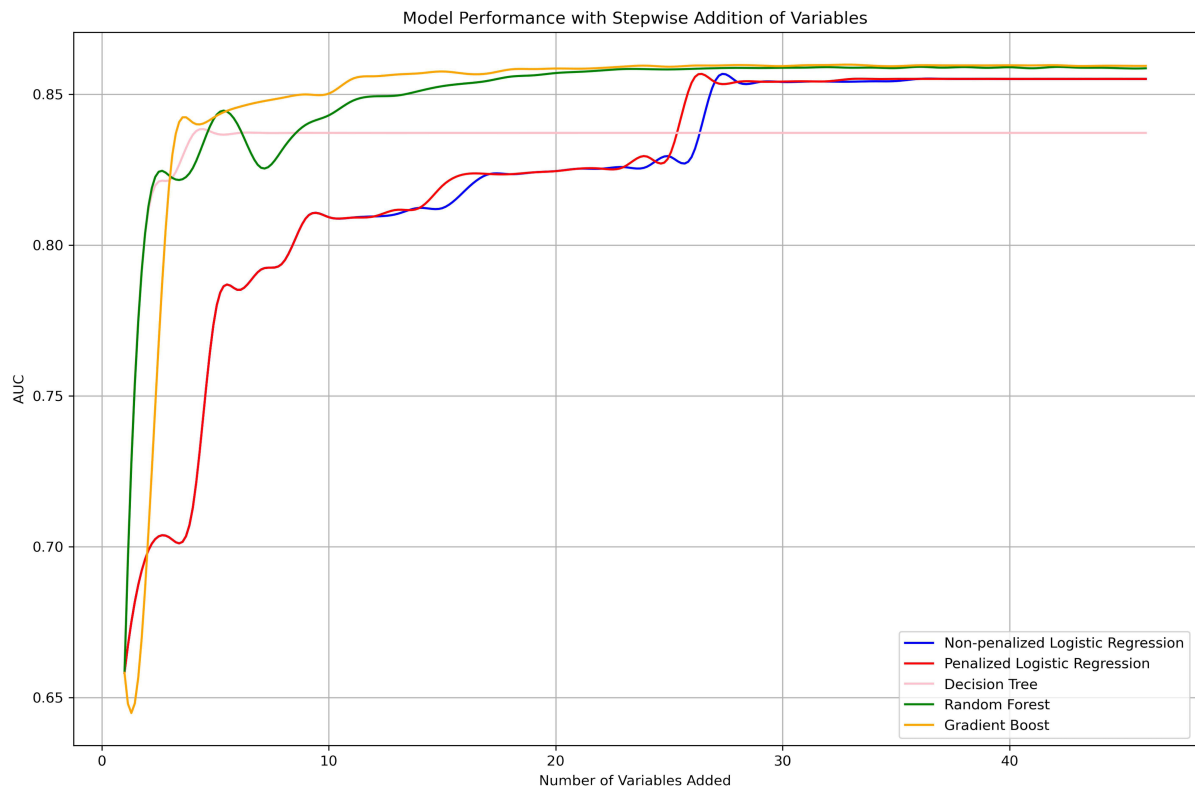
## Model Evaluation in Subset Analysis

Among the 641,042 participants included in our study population, a total of 28,372 participants (4.43%) had a previous cardiovascular event. We conducted a subset analysis by removing individuals who had a previous CVD event, ensuring the robustness of the models remained. As presented in Table 4, both penalized LR and GB did not show much difference in the performance metrics, whilst the AUC-ROC only decreased by 0.05. Similarly, in another subset analysis, a total of 110,716 participants (17.3%) were removed due to prior history of COPD. Very similar performance metrics were estimated in comparison to the inclusion of all features in the model.

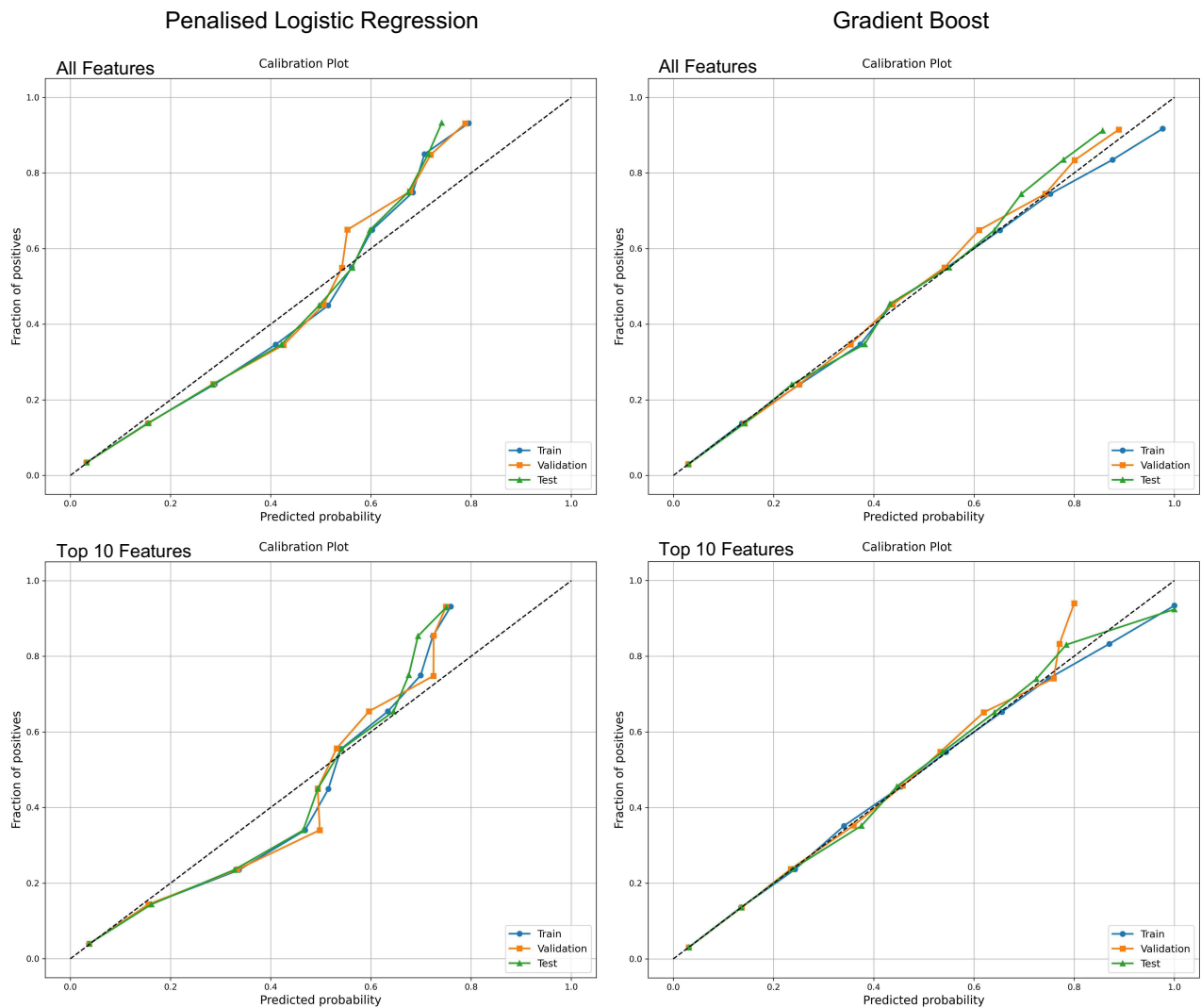
## Discussion

In this study, our analysis revealed that our various ML models produced similar results in the performance evaluation penalized LR model was the best and simplest model in terms of classification. From calibration curves, GB model was well-calibrated with relatively close predicted and observed probabilities closely followed by penalized LR. Hence, GB model was identified as the best model in terms of calibration.

Even though predictive models can determine the risk of an individual developing a specific condition and assist in the shared decision-making process, this does not guarantee an enhanced decision-making process.<sup>24</sup> The adoption and utilization of predictive models in clinical practice remains limited, and this has been evident in the relatively sparse coverage of ML models in clinical journals.<sup>25</sup> The lack of explainability on the way decisions are made in ML algorithms



**Figure 3** Model Performance with Stepwise Addition of Features – Top plot shows the AUC increase for the 5 models as the number of features are added sequentially from the feature ranking. Bottom plot shows the specificity of each model as features are added sequentially.



**Figure 4** Calibration Curves for Penalised Logistic Regression and Gradient Boost in the imbalanced dataset – Comparison of the calibration plots for penalised logistic regression (left) and gradient boost (right) models. Top row shows calibration curves of the models trained with all features, while the bottom row shows calibration curves of the models trained with the top 10 features. Perfect calibration indicated by dotted diagonal line.

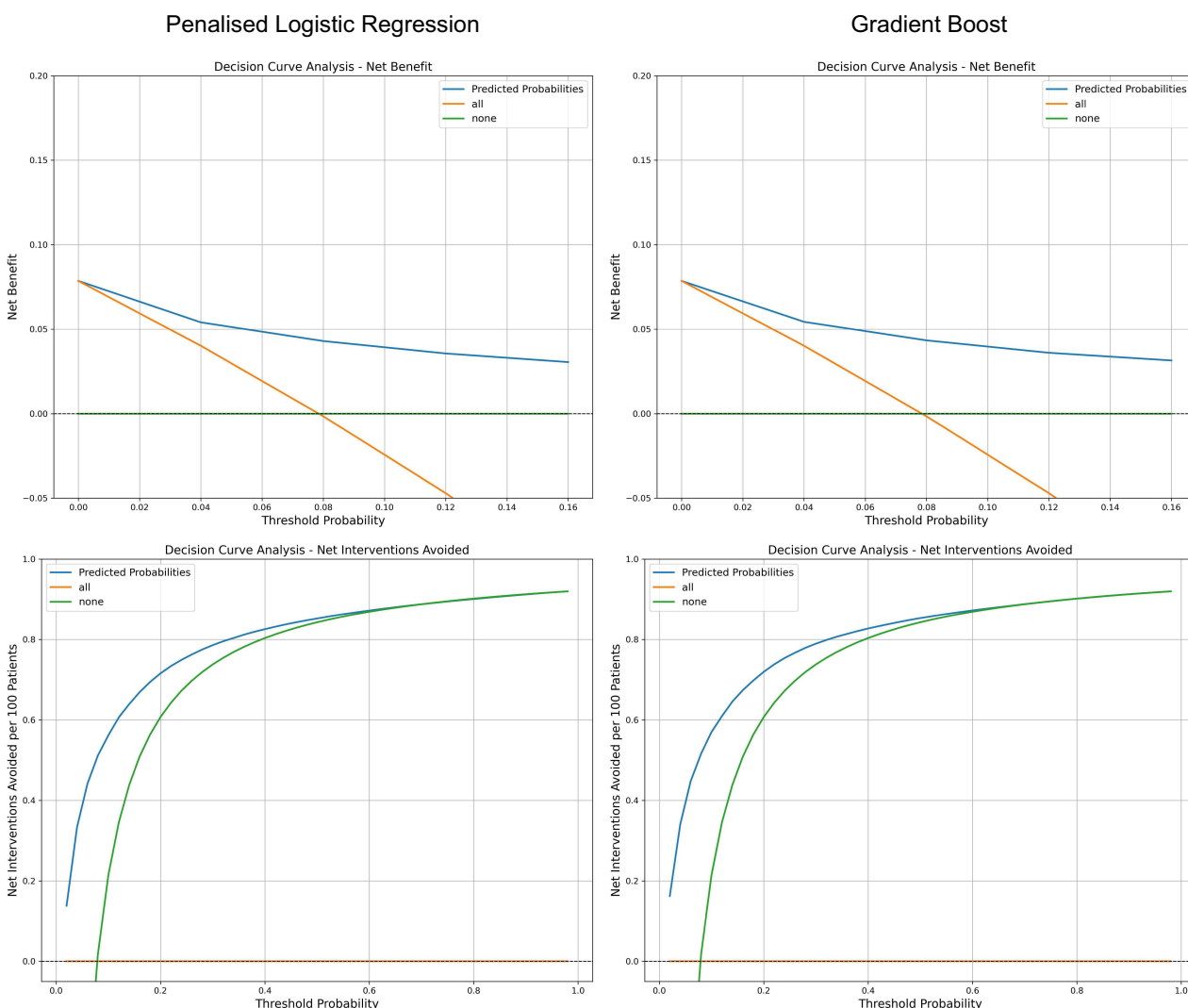
from the “black box” leads to a lack of trust and hindrance to utilize these models. Clinicians are more likely to use readily interpretable but less accurate standard models rather than more accurate ML models.

## Principle Findings – ML Models with and without Feature Selection

From both [Table 3](#) and [Supplementary 5](#), class weights resulted in no significant differences obtained from corresponding performance evaluation metrics.

Models trained on imbalanced data had very poor performance in accurately identifying the true positive cases. Sensitivity was estimated to be between 0.27 and 0.33 in models, despite high specificity values. Conversely, using class weights, sensitivity increased to 0.77 but there was reduction in specificity and precision. Our aim was to develop a risk prediction model to be used as a screening tool to rule out low-risk asthma individuals. This means a high specificity value would be preferable.

According to “SpPin and SnNout” rule,<sup>26</sup> SpPin suggests that when specificity is high, a positive test rules in the disease of interest and SnNout indicates that a negative test result would rule out the disease if sensitivity is high. The



**Figure 5** Clinical Utility for Penalised Logistic Regression and Gradient Boost – Net benefit shows trade-off between true positive and false positive predictions across a range of thresholds compared to treating all or none patients. Net reduction of interventions shows reduction in unnecessary treatments across a range of predicted probabilities compared to treating all or none patients.

authors describe why this rule can be problematic and suggest the calculation of likelihood ratios for a positive test result (LR+) and a negative test result (LR-) instead of SpPin and SnNout rules.

Using as an example penalized logistic regression estimates in Table 3, we estimated LR+ and LR- for imbalanced data. In this case, sensitivity was 0.27, and specificity was 0.99. From formulae available in Fisher et al paper, LR+ is equal to 27 which according to this paper would be conclusive to rule in patients. However, LR- is equal to 0.7, which based on the same paper would rarely be helpful to rule patients out. In addition, as sensitivity is equal to 0.27, this would mean as a result a relatively high percentage of false negatives and missing a lot of patients with the disease of interest.

Using weighted data for the same model (penalized logistic regression) with sensitivity 0.73 and specificity 0.82, LR+ is equal to 4.06 and LR- is equal to 0.33. Both estimates suggest a small effect on the probability of ruling patients in or out, respectively.

According to Tripod AI guidelines, decision curve analysis should be used to assess clinical utility of prediction models. Based on these guidelines, we implemented decision curve analysis as a method that is more likely to be informative compared to methods described in Fischer et al paper.

**Table 4** Performance of Penalised Logistic Regression and Gradient Boost Models in Patient Subgroups – Performance Metrics Presented Across Different Scenarios: Using All Features, Excluding Cases with Individuals with a Previous Cardiovascular Event in the Last 12 Months Since Index Date, and Excluding Individuals with a Prior History of COPD

Models	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
<b>All Features</b>						
Penalized Logistic Regression	0.93	0.27	0.99	0.64	0.38	0.85
Gradient Boost	0.93	0.32	0.98	0.63	0.42	0.86
<b>No Previous Cardiovascular Event in last 12 months</b>						
Penalized Logistic Regression	0.95	0.06	1.00	0.51	0.11	0.80
Gradient Boost	0.95	0.06	1.00	0.51	0.10	0.81
<b>No Prior History of COPD</b>						
Penalized Logistic Regression	0.94	0.25	0.99	0.64	0.36	0.85
Gradient Boost	0.95	0.30	0.99	0.63	0.41	0.86

In a real-world context, high specificity is desirable to reduce the number of false positives that would require unnecessary investigations. The choice of high specificity would have the drawback of lower sensitivity with the possibility of missing a few patients that would develop CVD.

False negatives are important for any screening method. By using class weights, sensitivity increased and simultaneously, there was a slight reduction in specificity. In this way, we were able to achieve balance between false positive and false negative rates which is also supported by our results from Decision Curve Analysis.

## Calibration Curves & Decision Curve Analysis

Even though a model has high AUC and specificity values, its results cannot be considered reliable for a comparison between predicted probabilities and observed rates unless the model is well calibrated.<sup>27</sup> When comparing calibration curves between different model types, imbalanced GB and penalized LR models emerged as the best-calibrated models.

DCA provides a measure of clinical utility and aids in translating the models into clinically meaningful outcomes which has been further emphasized by TRIPOD AI guidelines.<sup>21</sup> Both penalized LR and GB models at the outcome prevalence threshold resulted in net benefit with approximately 52%.

Therefore, both penalized LR and GB models are suitable for the risk prediction of CVD amongst individuals with asthma.

## Strengths, Limitations and Future Studies

Due to the large size of our study sample, minor variations in results can result in statistically significant results. For prediction models, predicted probabilities reflecting on the true likelihood of the event provides more utility in our study aiding in risk stratification and patient management.

Our analysis is based on a sufficiently large sample size for statistical analysis enhancing the validity of our findings. However, there may be selection bias as the dataset is derived from patients who visit general practices and have granted permission to contribute to CPRD.

Whilst CPRD is a comprehensive dataset within the UK, the findings may not be applicable to populations outside the UK due to variations in healthcare systems and disease prevalence.

Collaborating with other healthcare databases and potentially conducting prospective studies could enhance the utility and quality of our prediction model in healthcare settings.<sup>28</sup> In a future study, prospective validation of our models would be a valuable step forward as we would be able to assess the model's performance on patient outcomes.

Incorporating an external validation using independent datasets would enable for more rigorous examinations to be conducted, aiding in the verification of the generalizability of our models.

## Conclusion

In conclusion, this study explored various ML algorithms for CVD risk prediction. Imbalanced unweighted models are useful in accurate prediction of CVD event probabilities of high and low risk individuals. Whilst LR and GB models had similar results in terms of calibration and decision curve analysis, GB has a slightly better calibration. Penalized LR model might be preferable due to its simplicity, and GB model can be preferable to obtain further robustness and improved accuracy between the predicted and actual probabilities. The more advanced ML models like GB models are better at predicting cardiovascular risk probabilities, holding the potential to support healthcare environments.

Model interpretability refers to the transparency of the algorithm underlying each model. A logistic regression would be transparent by default as prediction probabilities can be derived from the corresponding predictor values multiplied by the corresponding predictor coefficients. Machine learning methods are regarded as less interpretable as the algorithms required are more convoluted and based on far more parameters compared to logistic regression models.

Surrogate models can be built to help with the interpretability of machine learning models. For example, logistic regression can be trained with the original machine learning model features to predict the result of classification of the machine learning models. However, interpretability will be still limited by the fact that the surrogate model will be simpler in structure compared to the original machine learning model but will be trained on the predicted outcome instead of the true (observed) outcome.

Penalized LR provides clear coefficients leading to simplicity in interpretability. The same type of model is less computationally intensive than GB, therefore faster training and prediction models were developed. GB models are effective in capturing complex interactions and non-linearities which LR models will not be able to capture without complex feature engineering. This type of model is more robust to outliers and therefore does not require normalization or standardization as it is less sensitive to outliers. Traditional models such as penalized LR cannot handle and struggle with data in text, images and other unstructured formats, whereas ML models can process these and produce structured inputs for use in hybrid modelling pipelines.

This study presents the potential to improve cardiovascular healthcare for asthma patients over the age of 50 through the implementation of predictive models. Challenges in implementation will need to be addressed through improved model transparency, robust clinical validation and interdisciplinary collaboration to increase the implementation and impact in healthcare.

Our findings can be implemented in primary and secondary care settings to identify patients with asthma at least 50 years old that have low risk of CVD within a year from asthma diagnosis. In this way, there will be a reduction of related investigations and treatments by 52% resulting in efficient care of patients with asthma.

## Data Sharing Statement

Data are available on request from the CPRD. Their provision requires the purchase of a license, and this license does not permit the authors to make them publicly available to all. This work used data from the version collected in September 2023 and has clearly specified the data selected within each Methods section. To allow identical data to be obtained by others, via the purchase of a license, the code lists will be provided upon request. Licenses are available from the CPRD (<http://www.cprd.com>): The Clinical Practice Research Datalink Group, The Medicines and Healthcare products Regulatory Agency, 10 South Colonnade, Canary Wharf, London E14 4PU.

**Author Contributions** All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Ethics Approval

CPRD has NHS Health Research Authority (HRA) Research Ethics Committee (REC) approval to allow the collection and release of anonymised primary care data for observational research [NHS HRA REC reference number: 05/MRE04/87]. Each year CPRD obtains Section 251 regulatory support through the HRA Confidentiality Advisory Group (CAG),

to enable patient identifiers, without accompanying clinical data, to flow from CPRD contributing GP practices in England to NHS Digital, for the purposes of data linkage [CAG reference number: 21/CAG/0008]. The protocol for this research was approved by CPRD's Research Data Governance (RDG) Process (protocol number: 23\_003184) and the approved protocol is available upon request. Linked pseudonymized data was provided for this study by CPRD. Data is linked by NHS Digital, the statutory trusted third party for linking data, using identifiable data held only by NHS Digital. Select general practices consent to this process at a practice level with individual patients having the right to opt-out.

This study is based in part on data from the Clinical Practice Research Datalink obtained under license from the UK Medicines and Healthcare products Regulatory Agency. The data is provided by patients and collected by the NHS as part of their care and support. The Office for National Statistics (ONS) was the provider of the ONS Data contained within the CPRD Data and maintains a Copyright © 2021, The Hospital Episode Statistics (HES) was the provider of HES-Admitted Patient Care and HES-Accident and Emergency databases contained within the CPRD Data and maintain a Copyright © 2021 and Copyright © 2020, respectively. Linked data were re-used with the permission of The Health & Social Care Information Centre, all rights reserved. The interpretation and conclusions contained in this study are those of the author/s alone.

## Author Contributions

JKQ and CK conceptualized the study and designed the protocol. CK and JKQ contributed to the development of the code lists that defined the study variables. JB, JKQ and CK contributed to the methodology. JB, VAN and CK were responsible for data curation and management. JB and CK were responsible for formal analysis. JB, JKQ and CK wrote the original draft of the manuscript. JB, VAN and CK were responsible for data visualization. JB, VAN, JKQ and CK reviewed the manuscript. All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This research was supported by the NIHR Imperial Biomedical Research Centre (BRC).

## Disclosure

Professor Jennifer Quint reports grants from HDR UK, during the conduct of the study; grants and/or personal fees from AZ, Chiesi, Sanofi, NIHR, outside the submitted work. The authors report no other conflicts of interest in this work.

## References

1. Health Intelligence Team B. *BHF England CVD Factsheet*. 2024.
2. British Lung Foundation. Asthma statistics | British Lung Foundation. 2024 Available from: <https://statistics.blf.org.uk/asthma>. Accessed August 14, 2024.
3. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One*. 2017;12(4):e0174944. doi:10.1371/journal.pone.0174944
4. MAEFMD B, Naing L, Johar S, et al. Evaluation of cardiovascular diseases risk calculators for CVDs prevention and management: scoping review. *BMC Public Health*. 2022;22(1):1–11.
5. Ridker PM, Danielson E, Fonseca FAH, et al. Rosuvastatin to prevent vascular events in men and women with elevated c-reactive protein. *N Engl J Med*. 2008;359(21):2195–2207. doi:10.1056/NEJMoa0807646
6. Abdar M, Yen NY, Hung JCS. Improving the diagnosis of liver disease using multilayer perceptron neural network and boosted decision trees. *J Med Biol Eng*. 2018;38(6):953–965. doi:10.1007/s40846-017-0360-z
7. Khozimeh F, Alizadehsani R, Roshanzamir M, Khosravi A, Layegh P, Nahavandi S. An expert system for selecting wart treatment method. *Comput Biol Med*. 2017;81:167–175. doi:10.1016/j.compbiomed.2017.01.001
8. CPRD. CPRD Aurum September 2023 dataset | CPRD Available from: <https://www.cprd.com/doi/cprd-aurum-september-2023-dataset>. Accessed August 14, 2024.
9. Wolf A, Dedman D, Campbell J, et al. Data resource profile: clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol*. 2019;48(6):1740–1740g. doi:10.1093/ije/dyz034
10. Nissen F, Morales DR, Mullerova H, Smeeth L, Douglas IJ, Quint JK. Validation of asthma recording in the Clinical Practice Research Datalink (CPRD). *BMJ Open*. 7(8):1–9.

11. Kallis C, Maslova E, Morgan AD, et al. Recent trends in asthma diagnosis, preschool wheeze diagnosis and asthma exacerbations in English children and adolescents: a SABINA Jr study. *Thorax*. 2023;78(12):1175–1180. doi:10.1136/thorax-2022-219757
12. Kallis C, Calvo RA, Schuller B, Quint JK. Development of an asthma exacerbation risk prediction model for conversational use by adults in England. *Pragmat Obs Res*. 2023;14:111–125. doi:10.2147/POR.S424098
13. Varoquaux G, Pedregosa F, Gramfort A, et al. LogisticRegression — scikit-learn 1.5.1 documentation. Available from: [https://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html). Accessed August 14, 2024.
14. Louppe G, Prettenhofer P, Holt B, et al. DecisionTreeClassifier — scikit-learn 1.5.1 documentation. 2024 Available from: <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html>. Accessed July 30, 2024.
15. Louppe G, Holt B, Arnaud J, Hedayati F. RandomForestClassifier — scikit-learn 1.5.1 documentation 2024 Available from: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html>. Accessed July 30, 2024.
16. Prettenhofer P, White S, Louppe G, Olivetti E, Joly A, Schreiber J. GradientBoostingClassifier — scikit-learn 1.5.1 documentation [Internet]. 2024 Available from: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>. Accessed July 30, 2024.
17. Mueller A, Kumar M. compute\_sample\_weight — scikit-learn 1.5.1 documentation Available from: [https://scikit-learn.org/stable/modules/generated/sklearn.utils.class\\_weight.compute\\_sample\\_weight.html](https://scikit-learn.org/stable/modules/generated/sklearn.utils.class_weight.compute_sample_weight.html). Accessed July 30, 2024.
18. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5(9):1315–1316. doi:10.1097/JTO.0b013e3181ec173d
19. Hosmer DW, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Third Edition. Applied Logistic Regression: Third ed.
20. Hicks SA, Strümke I, Thambawita V, et al. On evaluation metrics for medical applications of artificial intelligence. *Sci Rep*. 2022;12(1) doi:10.1038/s41598-022-09954-8
21. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. Available from: <http://www.bmj.com/>. Accessed July 30, 2024.
22. Bavo De Cock Campo. Introduction to the CalibrationCurves package. 2024 Available from: <https://cran.r-project.org/web/packages/CalibrationCurves/vignettes/CalibrationCurves.html>. Accessed July 30, 2024.
23. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and interpreting decision curve analysis: a guide for investigators. *Eur Urol*. 2018;74(6):796–804. doi:10.1016/j.eururo.2018.08.038
24. Hassan N, Slight R, Morgan G, et al. Road map for clinicians to develop and evaluate AI predictive models to inform clinical decision-making. *BMJ Health Care Inform*. 2023;30(1):100784.
25. Kwan JL, Lo L, Ferguson J, et al. Computerised clinical decision support systems and absolute improvements in care: meta-analysis of controlled clinical trials. *BMJ*. 370;3216.
26. Fischer BG, Evans AT. SpPin and SnNout are not enough. it's time to fully embrace likelihood ratios and probabilistic reasoning to achieve diagnostic excellence. *J Gen Intern Med*. 2023;38(9):2202–2204. doi:10.1007/s11606-023-08177-5
27. Razavizadeh NT, Salari M, Jafari M, Sabaghian E, Ghavami V. Comparison of two methods, gradient boosting and extreme gradient boosting to pre-dict survival in covid-19 data. *J Biostat Epidemiol*. 2023;9.
28. Wang SCY, Nickel G, Venkatesh KP, Raza MM, Kvedar JC. AI-based diabetes care: risk prediction models and implementation concerns. *npj Digital Medicine* 2024. 7;1.<https://www.nature.com/articles/s41746-024-01034-7>.

## Pragmatic and Observational Research

### Publish your work in this journal

Pragmatic and Observational Research is an international, peer-reviewed, open access journal that publishes data from studies designed to reflect more closely medical interventions in real-world clinical practice compared with classical randomized controlled trials (RCTs). The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/pragmatic-and-observational-research-journal>

**Dovepress**  
Taylor & Francis Group