IMPERIAL

Investigating Demographic, Psychosocial, and Biological Factors Predictive of Prolonged Benzodiazepine Prescription in a UK Biobank Primary Care Cohort: A Machine Learning Approach.

TDS Group 3

Dan Rose, Wanqing Fang, Yilin Chen, Titouan Guillou, Berin Pinar
30/04/2025

Outline

Background

Benzodiazepine prescription and long-term use in the UK

- Over 1.4M benzodiazepine (BZD) prescriptions are dispensed annually in the UK , most commonly for anxiety and sleep disorders.
- **In England 35% of recipients are long-term* users, amounting to an estimated 300,000 individuals.**
- Studies investigating determinants of recurrent BZD prescription cite sex, socioeconomic status, and comorbid disease as strong predictors.
- Widespread controversy exists surrounding BZD prescription due to side-effect and efficacy concerns.

Research Gap:
- No studies exist that predict **long-term** prescription nor **stratified by sex**.
- Contribution of **biomarkers** and/or **anxiety status** to outcome not explored.
- Kinney et al. (2023) echoes the need of a model able to predict prolonged BZD use.

** ≥12 months*

Our research question:

Can **demographic, environmental, psychological,** and **biological factors** be utilized to **accurately predict** long-term **BZD** prescription?

Aims

Among patients who have received a GP prescription we aim to determine:

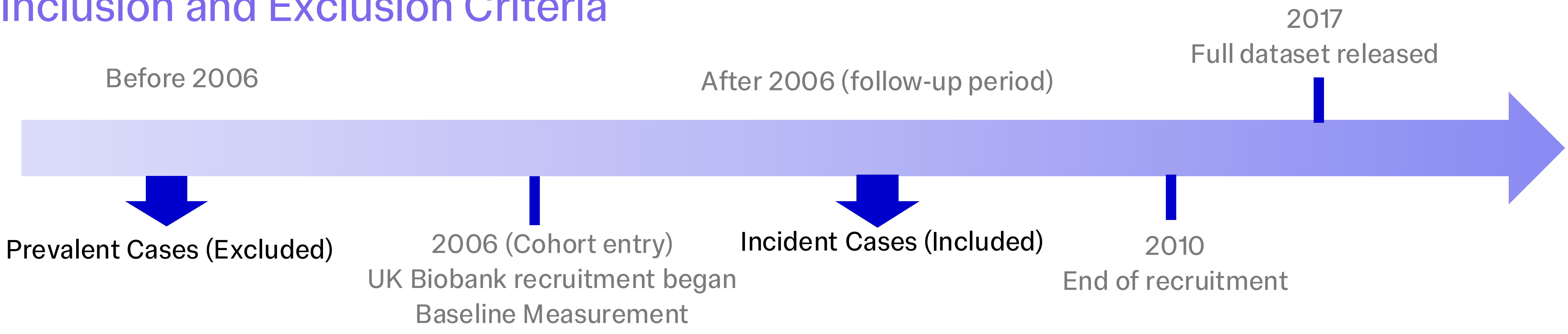
I. **What demographic, environmental, and behavioral factors are connotated with receiving long-term BZD prescription?**

II. **Does stratifying individuals based on their anxiety status improve the predictive power of the models?**

III. **Can biomarkers be used to predict prolonged BZD prescription, are there predictive gains when used in conjunction with above models?**

Data Processing

Inclusion and Exclusion Criteria



Cases:

- Individuals receiving ≥2 BZD prescriptions within 365 days

Controls:

- Individuals who received 1≤ BZD prescription in a year
- Individuals who had >1 prescriptions over 365 days apart
- Those never had a BZD prescription

Selection Criteria

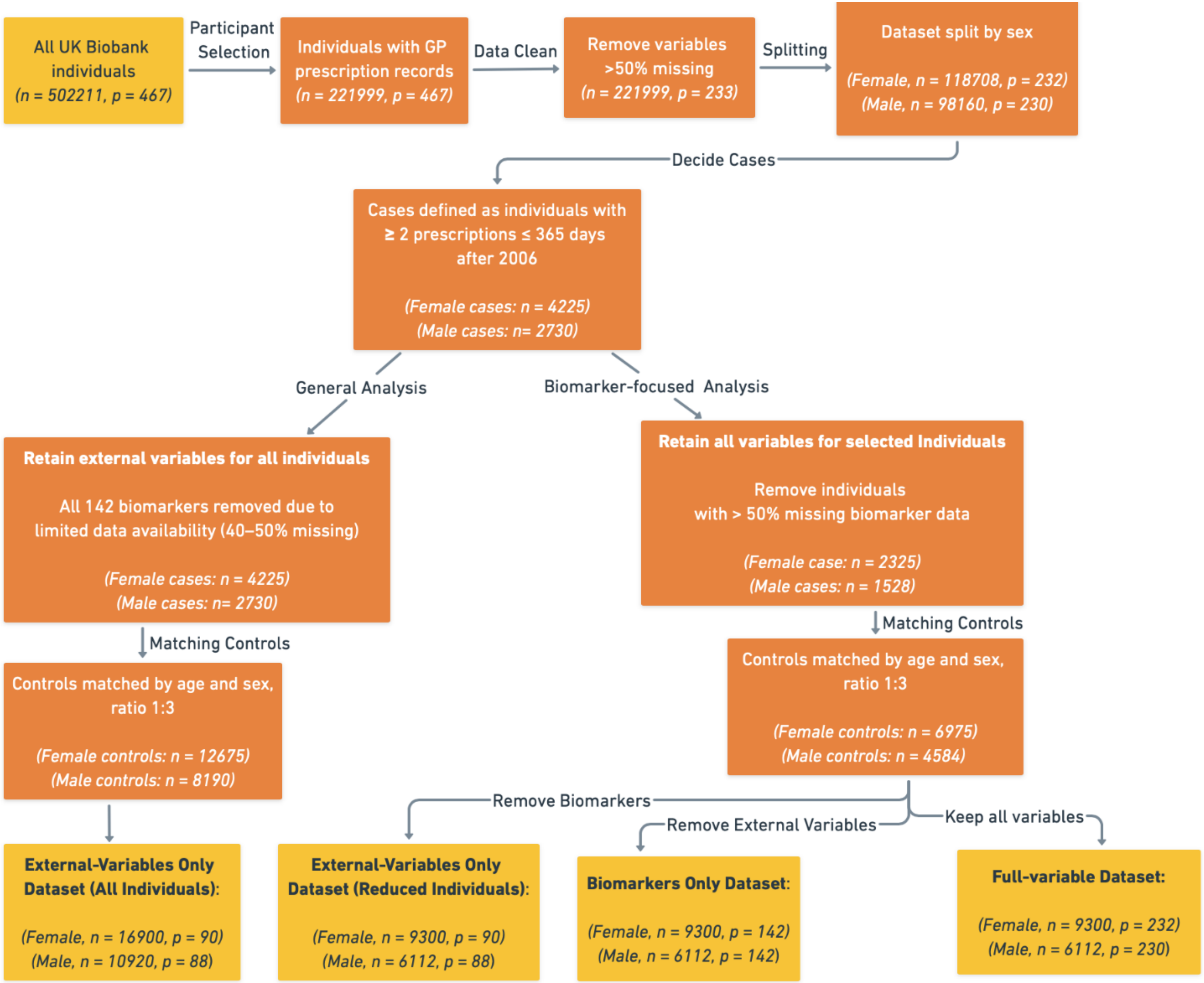
- UK Biobank participants with linked GP prescription data
- Retain only individuals with baseline measurements from the UK Biobank dataset

Exclusion Criteria

- Individuals without a history of GP prescription
- Individuals with ≥2 benzodiazepine prescriptions before 2006 (prevalent cases)

Data Processing Schematic

Pipeline



Data Processing

Summary of variables used and missingness in the Dataset

Missingness Level	Variable Group	Example Variables
Moderate (20%-30%)	Socioeconomic Factors	Crime score, living environment score
	Clinical Factors	Cancer diagnosis by doctor
Low (10%-20%)	External Factors	Greenspace percentage, employment score, weekly usage phone
Very Low (<10%)	Lifestyle Factors	BMI, weight, alcohol drinking status, current smoking, sleeplessness, health score, education score, family-friends visit
	Environmental Factors	Air pollution (nitrogen dioxide, PM2.5, PM10), noise pollution levels
	Working mode	Frequency of work not at home, number of hours working per week (after imputed unemployed as 0)
	Mental Health	Anxiety-related variables, mood swings, guilty feelings
No Missing (0%)	Demographics	Age, age at recruitment, genetic sex, ethnic background

Methodology

Modelling approaches

1. Stratified gender analysis

- Regression models (LASSO and Elastic Net) were applied. Stability selection was used after to identify stable predictors.
- Ensemble models (Random Forest and XGBoost) were performed

2. Anxiety subset analysis

- LASSO models were applied to each subgroup, followed by stability selection to identify stable predictors. Odd ratios were obtained using logistic regression on stably selected variables
- Random Forest models were performed separately

3. Biomarker focused analysis (using population with available biomarker measurements)

- Stratified by sex.
- LASSO and Random Forest on : (i) biomarkers only, (ii) non-biomarkers (external variables) only, and (iii) all variables combined

Descriptive Analysis

	Female (n = 16,900)				Male (n = 10,920)			
	Case	Control	p-value	Missing(%)	Case	Control	p-value	Missing(%)
n	4,225	12,675			2,730	8,190		
Demographics								
Age	75.00 [67.00, 80.00]	75.00 [67.00, 80.00]	1	0	76.00 [68.00, 81.00]	76.00 [68.00, 81.00]	1	0
Ethnic background = White	4,021 (95.6)	12,042 (95.4)	0.613	0.4	2,600 (95.8)	7,772 (95.4)	0.411	0.5
Owner or rent			<0.001	4.9			<0.001	5.5
Own	3,544 (88.6)	11,305 (93.7)			2,228 (87.9)	7,238 (93.0)		
Rent	435 (10.9)	668 (5.5)			285 (11.2)	471 (6.1)		
Others	20 (0.5)	96 (0.8)			22 (0.9)	74 (1.0)		
Current employment status = Unemployed	2,218 (53.1)	5,855 (46.6)	<0.001	1	1,375 (51.0)	3,348 (41.3)	<0.001	1
External variables								
Home area population density			0.005	1			0.02	1.4
Urban	3,613 (86.5)	1,0619 (84.6)			2,328 (86.6)	6,822 (84.5)		
Small Town	328 (7.9)	1,080 (8.6)			214 (8.0)	720 (8.9)		
Village	153 (3.7)	558 (4.4)			91 (3.4)	343 (4.2)		
Rural	34 (0.8)	80 (0.6)			19 (0.7)	40 (0.5)		
Hamlet and Isolated Dwelling	50 (1.2)	215 (1.7)			37 (1.4)	153 (1.9)		
Education score	6.79 [1.15, 19.25]	7.72 [1.93, 19.80]	<0.001	2.4	7.58 [1.58, 22.05]	8.03 [2.02, 20.80]	0.268	2.9
Income score	0.12 [0.06, 0.33]	0.11 [0.05, 0.30]	<0.001	2.4	0.11 [0.06, 0.28]	0.11 [0.05, 0.30]	0.524	2.8
Employment score	0.07 [0.05, 0.12]	0.08 [0.05, 0.13]	<0.001	16.2	0.07 [0.05, 0.12]	0.08 [0.05, 0.14]	<0.001	15.1
Crime score	0.01 [-0.5, 0.52]	-0.05 [-0.58, 0.47]	<0.001	22.9	0.03 [-0.5, 0.53]	-0.01 [-0.53, 0.5]	0.043	22
Housing score	18.42 [11.05, 27.25]	17.38 [10.15, 25.78]	<0.001	2.4	18.01 [10.66, 26.44]	17.37 [10.13, 25.84]	0.005	2.8
IMD score	13.70 [7.60, 25.16]	12.55 [7.00, 22.19]	<0.001	2.4	13.70 [7.40, 26.37]	12.65 [7.23, 23.09]	<0.001	2.8
Living environment score	15.38 [7.39, 27.03]	14.37 [7.18, 25.79]	0.014	22.7	15.75 [7.75, 26.97]	14.10 [7.16, 26.14]	0.003	21.8
(Mental) Health Variable								
Anxiety / anxious individuals	2,747 (65.7)	5,060 (40.3)	<0.001	1	1,320 (48.9)	2,114 (26.0)	<0.001	0.9
Health score	-0.05 [-0.6, 0.49]	-0.01 [-0.57, 0.6]	<0.001	2.5	-0.04 [-0.6, 0.54]	0.03 [-0.53, 0.65]	<0.001	3
Overall health rating			<0.001	0.6			<0.001	0.6
Poor	387 (9.2)	459 (3.6)			376 (13.9)	409 (5.0)		
Fair	1,191 (28.4)	2,417 (19.2)			812 (30.0)	1,855 (22.8)		
Good	2,220 (52.9)	7,491 (59.5)			1,257 (46.5)	4,620 (56.7)		
Excellent	396 (9.4)	2,231 (17.7)			259 (9.6)	1,266 (15.5)		
Worrier / anxious feelings = Yes	3,072 (74.7)	7,735 (62.8)	<0.001	2.8	1,565 (58.6)	3,697 (46.6)	<0.001	2.8
Tense / 'highly strung' = Yes	1,324 (33.0)	2,172 (17.8)	<0.001	4	688 (26.3)	1,188 (15.0)	<0.001	3.7
Irritability = Yes	1,340 (33.3)	3,083 (25.5)	<0.001	4.8	1,054 (40.9)	2,333 (30.0)	<0.001	5.1

N(%) used for categorical variables, N(IQR) for continuous variables.

Descriptive Analysis
Prescription Data

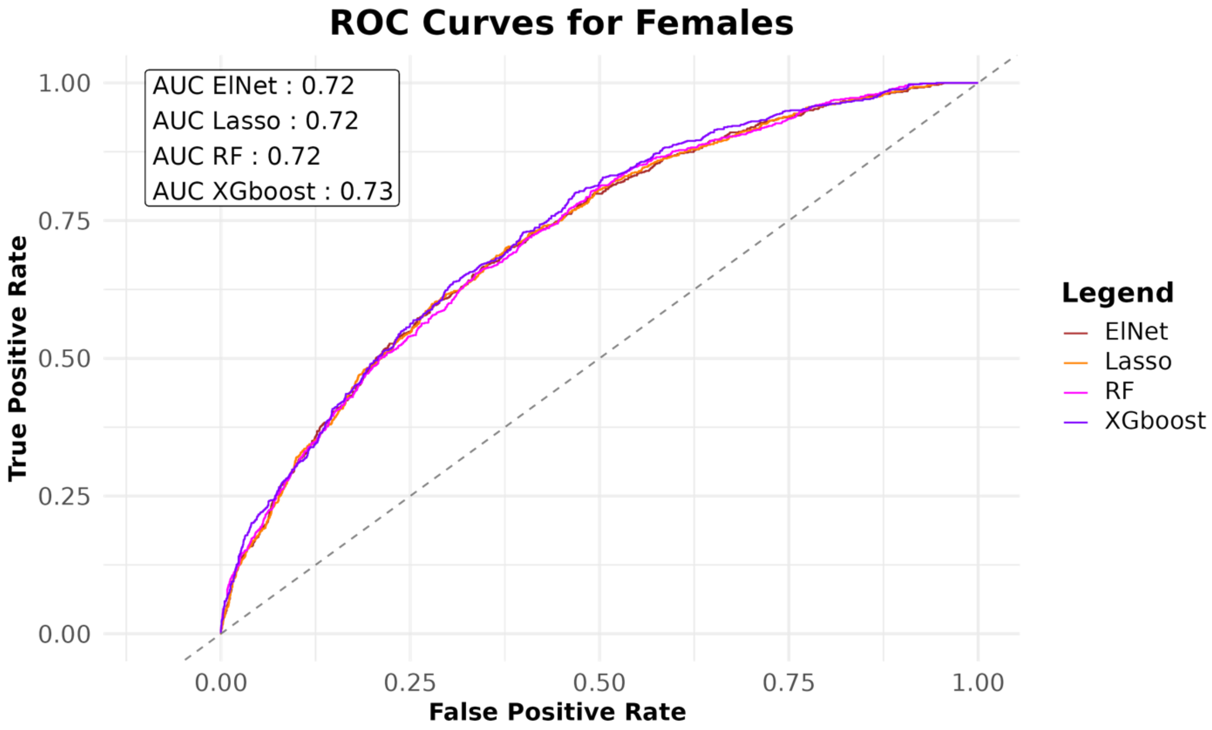
	Female			Male		
	Case (%)	Control (%)	Total	Case (%)	Control (%)	Total
n	4225	12675	16900	2730	8190	10920
Number of prescriptions						
0	0	11680 (92.1)	11680	0	7712 (94.2)	7712
1	0	846 (6.7)	846	0	426 (5.2)	426
2	1252 (29.6)	123 (1.0)	1375	929 (34.0)	43 (0.5)	972
3+	2973 (70.4)	26 (0.2)	2999	1801 (66.0)	9 (0.1)	1810
BZD prescription						
Drug name						
Diazepam	18759 (51.3)	951 (80.8)	19710	12486 (46.3)	391 (72.6)	12877
Temazepam	7442 (20.3)	178 (15.1)	7620	5235 (19.4)	94 (17.4)	5329
Clonazepam	5070 (13.9)	9 (0.8)	5079	5330 (19.7)	5 (0.9)	5335
Lorazepam	2976 (8.1)	13 (1.1)	2989	2078 (7.7)	12 (2.2)	2090
Others	2343 (6.4)	26 (2.2)	2369	1872 (6.9)	37 (6.9)	1909
Total prescription	36590	1177	37767	27001	539	27540
Average prescription (SD)	8.66 (17.9)	0.09 (0.35)	2.23 (9.7)	9.89 (24.5)	0.07 (0.3)	2.52 (12.9)

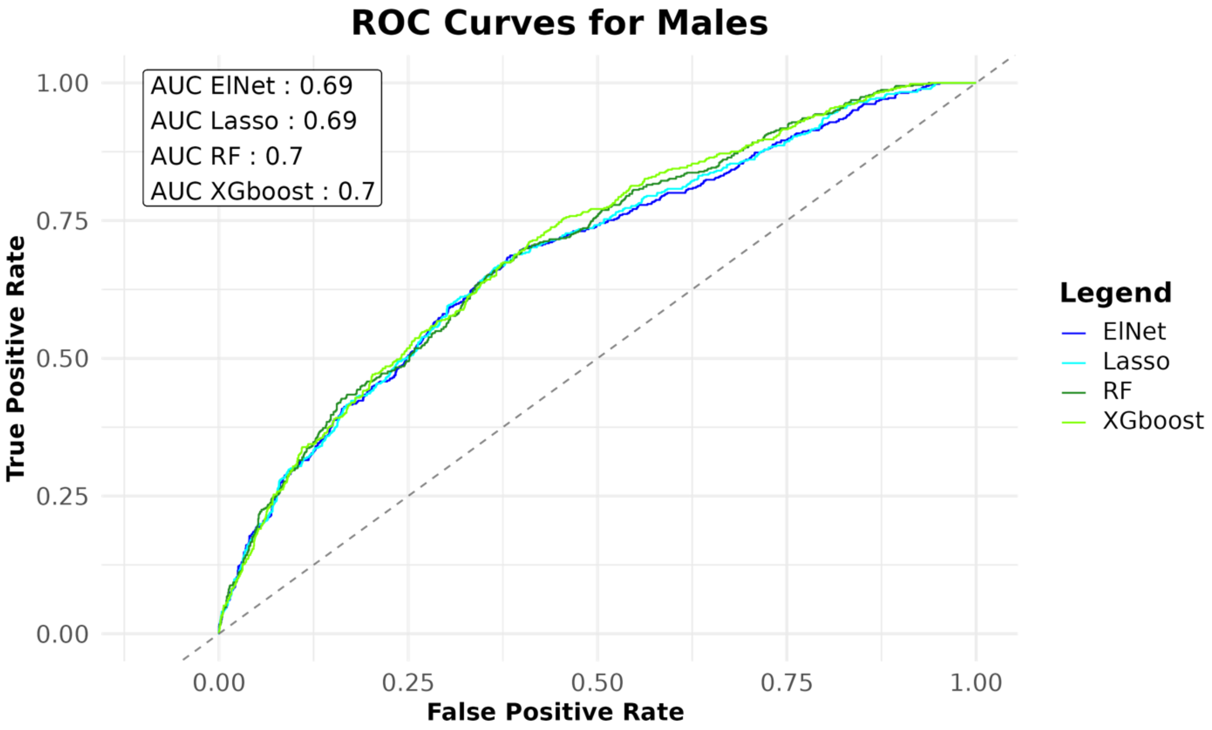
N(%) used for categorical variables, Mean(SD) for continuous variables.

GP prescription period: 03/01/2006 to 12/07/2017

Results

1. Predictive Modelling: A Gender Split Analysis





→ Overall similar performances of all models, slightly better for females.

Results

1. Predictive Modelling: A Gender Split Analysis

Sex	Models	AUC	Accuracy	Sensitivity	Specificity	F1-Score
Female	Lasso	0.72	**0.76**	0.10	**0.98**	0.17
	Elastic Net	0.72	**0.76**	0.11	**0.98**	0.19
	Random Forest	0.72	0.68	**0.58**	0.71	**0.61**
	XGBoost	**0.73**	**0.76**	0.10	**0.98**	0.17
Male	Lasso	0.69	**0.76**	0.08	**0.98**	0.14
	Elastic Net	0.69	**0.76**	0.10	**0.98**	0.17
	Random Forest	**0.70**	0.69	**0.49**	0.75	**0.55**
	XGBoost	**0.70**	0.75	0.11	0.97	0.19

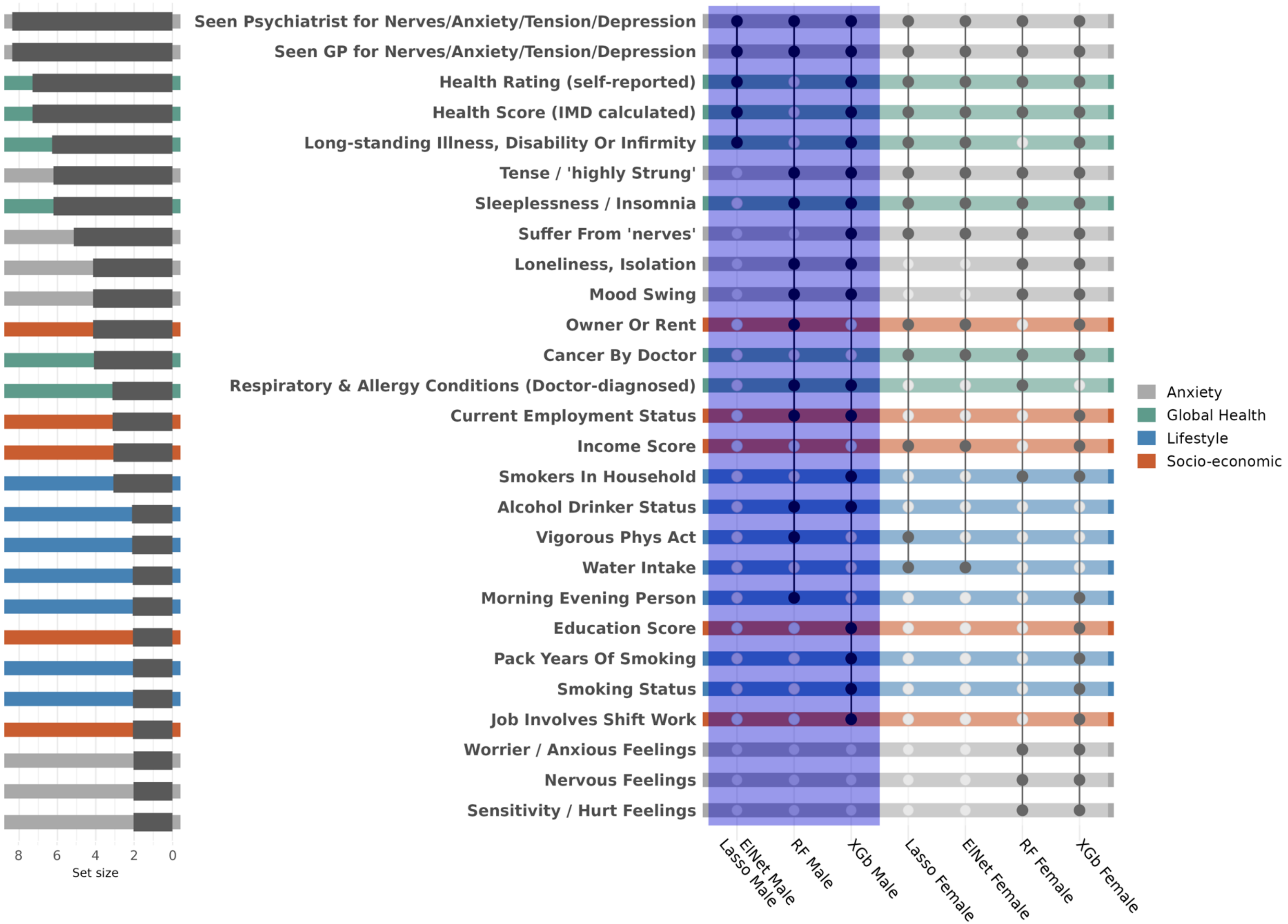
Results

1. Predictive Modelling: A Gender Split Analysis

**Features selected using :*

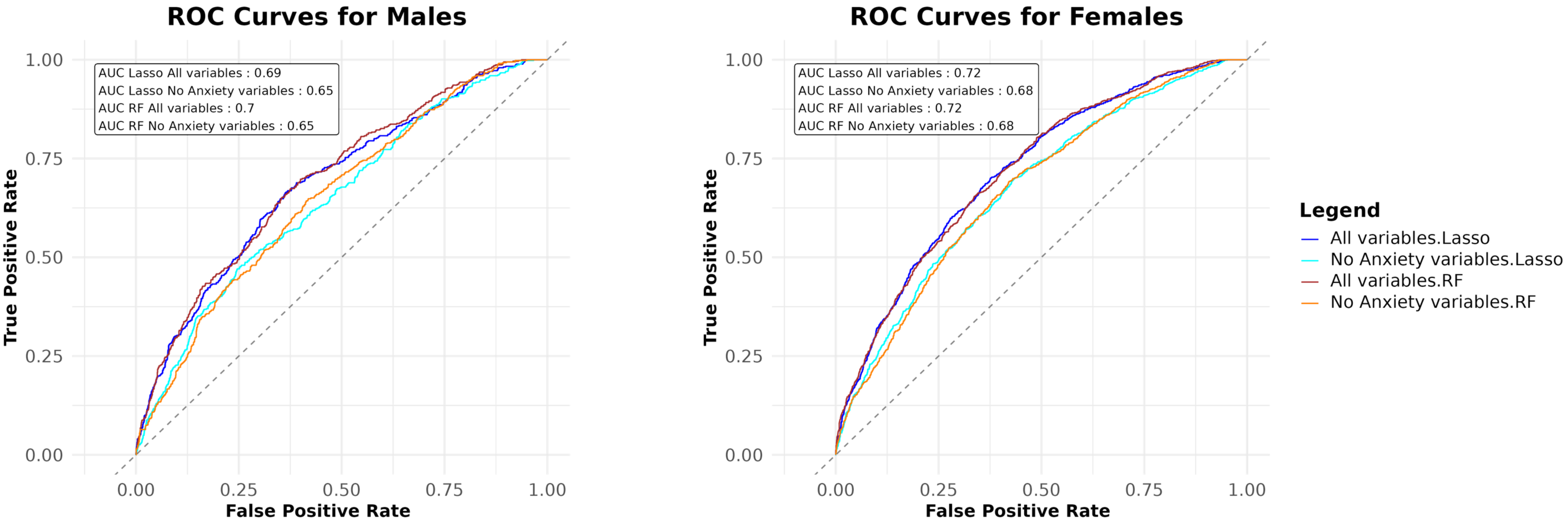
- *Stability Selection for **LASSO** and **elastic net***
- *Positive permutation score for **RF***
- *Gain importance larger than the 3rd quartile value of all the gains for **XGBoost***

Upset Plot for the different models *



Results

1. Predictive Modelling: Gender Analysis **without** anxiety related variables



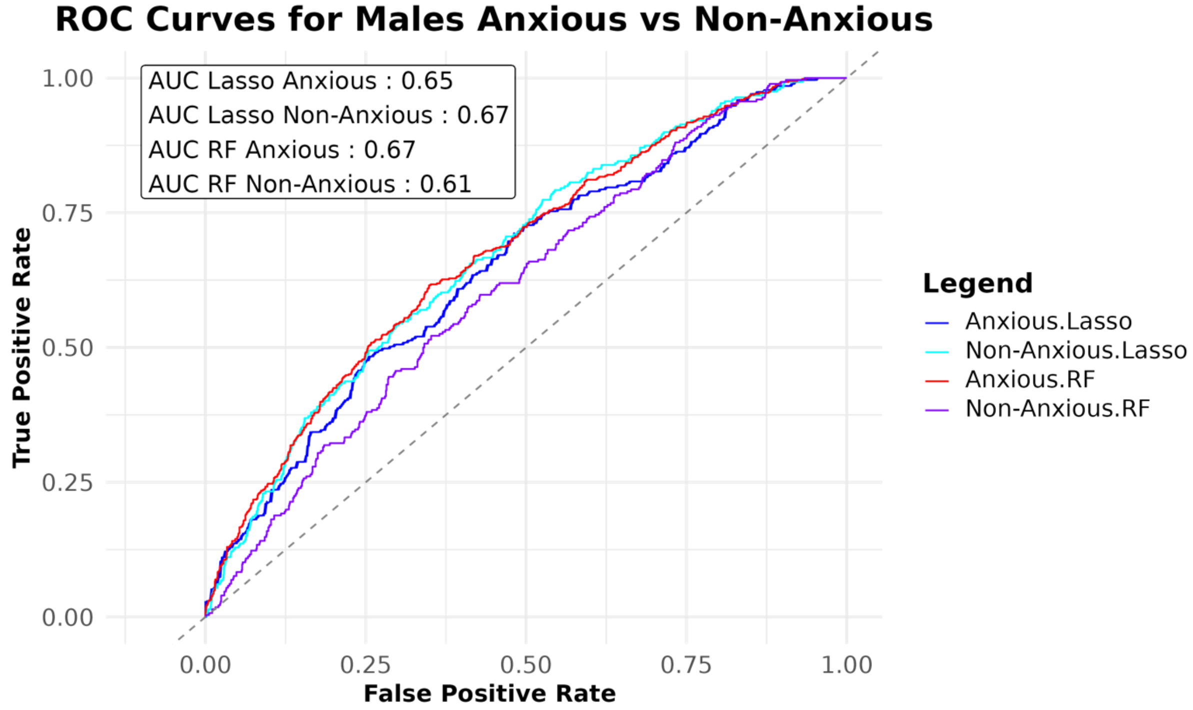
→ Slight loss of power when the anxiety related variables are removed from the prediction features

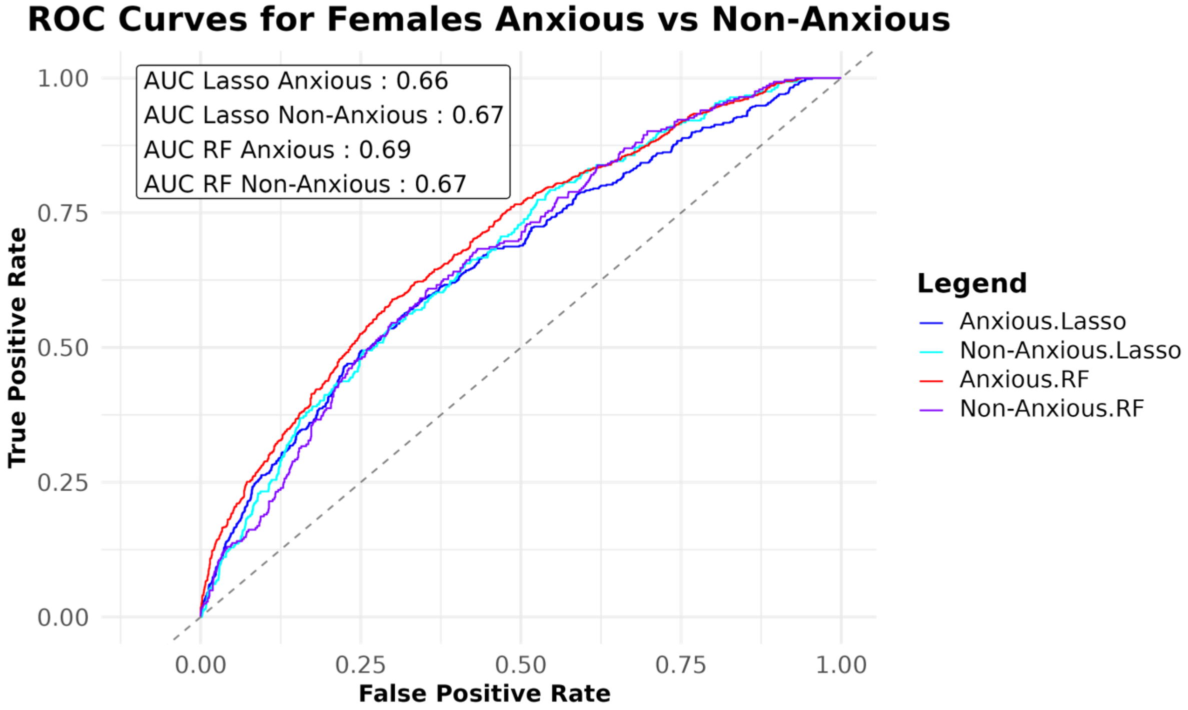
Results

2. Predictive Modelling: Determining Predictive Power in Anxious Individuals

Anxious indiviudals : *Have ever seen doctor and/or psychiatrist (GP) for nerves, anxiety, tension or depression*

Male anxious = 3441, Male non-anxious = 7479, Female anxious = 7836, Female non-anxious = 9064

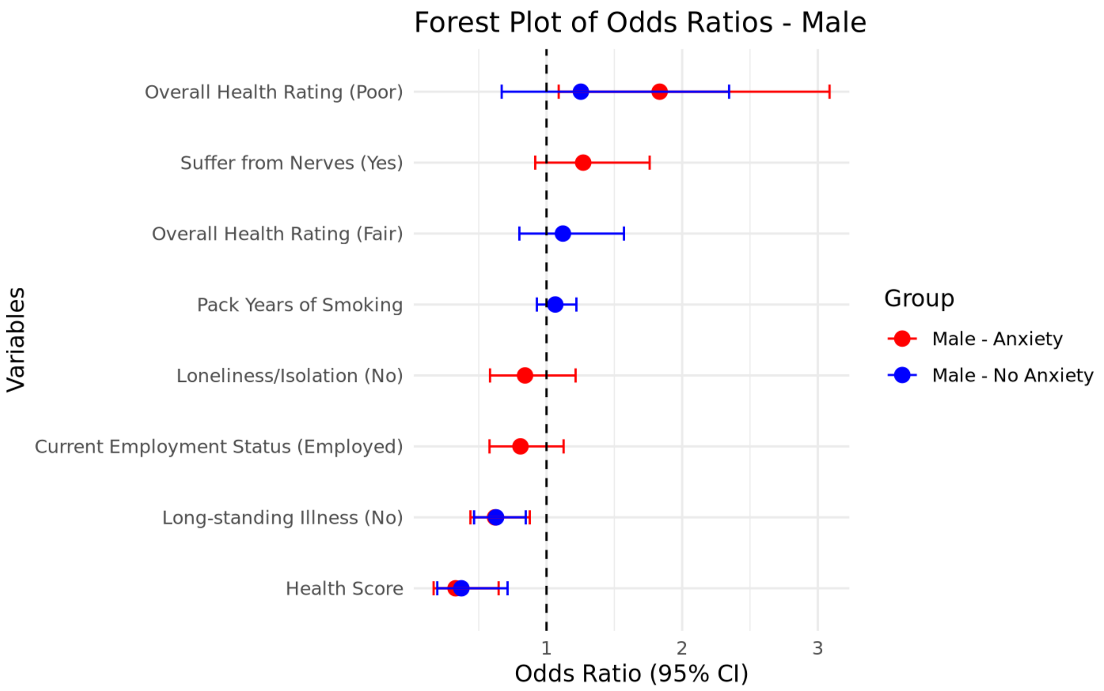


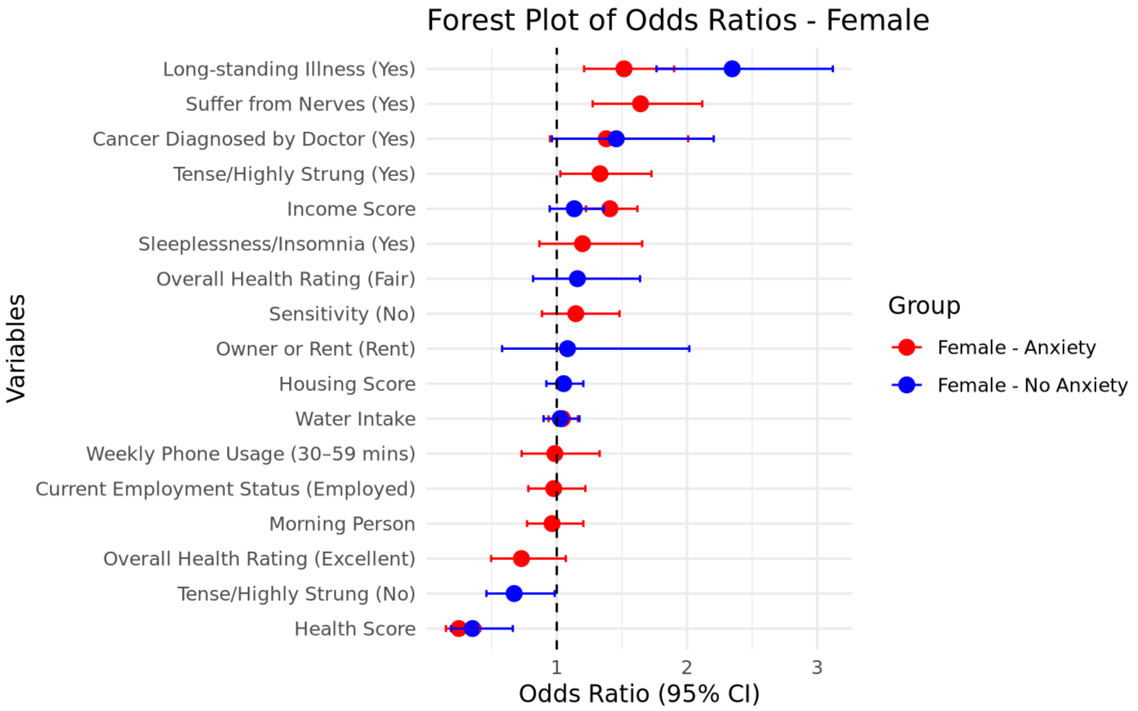


- LASSO showed better performance for non-anxious group, but RF performed better for anxious group in both sexes
- Overall range of model performances is relatively narrow (AUCs 0.61-0.69)

Results

2. Logistic Regression for Stably Selected Variables



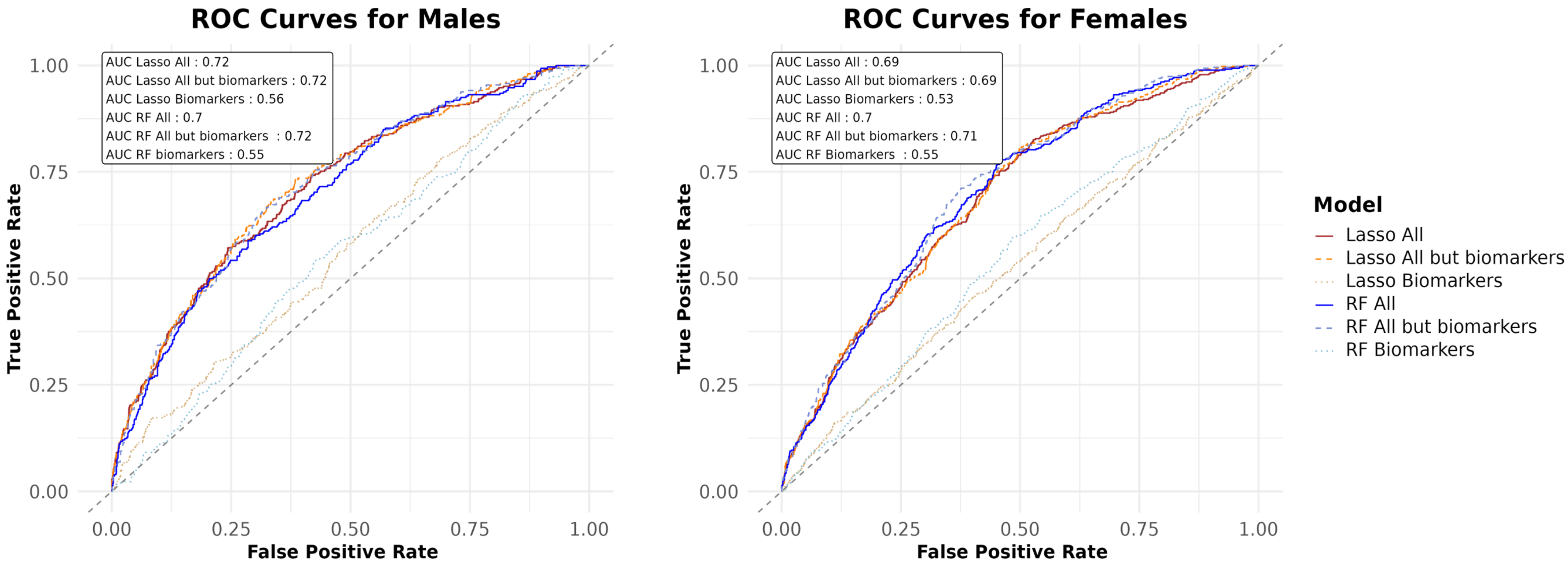


- Male: Physical health variables have significant associations. Poor health rating increased odds of repeat BZD prescribing by ~80% (anx), while absence of long-standing illness reduced odds by ~40% (both).
- Female: Mental health variables are strong and statistically significant predictors: being tense increases the odds by 30% (anx)

See appendix for Odds Ratios (95% CI, p-value)

Results

3 - Predictive Modelling: A Biomarker Perspective



- Biomarkers have limited predictive power for predicting prolonged BZD prescription.

Discussion
Insights

Model Performance

- All models in the sex stratified analysis achieved moderate to good predictive performance, however, the performance dropped when anxiety variables were removed
- Predictive performance did not substantially differ based on anxiety status
- Biomarkers measured at baseline provided only marginal discrimination and did not enhance the previous models

Conclusion

- Broader clinical, demographic, behavioural and socio-economic factors influence long-term BZD use across both sexes, regardless of anxiety status
- Predictive models should include wide variety of predictors to facilitate:
  - Targeted interventions for people at risk of prolonged BZD use
  - Alternative therapeutic approaches
  - Enhanced patient monitoring

Study Limitations

Ambiguous case/anxiety definitions

- Individuals with only one long-term prescription in a year may be misclassified as non-cases
- Individuals with two or more short-term prescriptions may be misclassified as long-term users

Lack of external validation

Model was only validated internally which tends to overestimate prediction

UK Biobank cohort

Elder and pre-dominantly white population of UK Biobank limits the generalisability of our findings to diverse populations

Future Considerations

External validation

- Predictive algorithms utilising diverse populations beyond the elderly UK Biobank cohort
- Explore differences in BZD prescribing patterns across different age groups and countries

Address ambiguity of case definitions

- Incorporating detailed prescription metrics (dosage, duration, reason for prescription)
- Improving accurate identification of true long-term BZD users

Subclassification and alternative models

- Combined approaches to identify individuals at increased risk of BZD related drug misuse or addiction → moving towards personalised medicine

Impact analyses

- Ensuring algorithms are not only effective from a predictive standpoint but that they also improve outcomes from a clinical and economic perspective

Thank You

Questions are welcomed

References

1. Public Health England. Prescribed medicines review: summary. London: GOV.UK; 2019 [cited 2025 Apr 24]. Available from: https://www.gov.uk/government/publications/prescribed-medicines-review-report/prescribed-medicines-review-summary
2. Davies J, Read J. A systematic review into the incidence, severity and duration of antidepressant withdrawal effects: Are guidelines evidence-based? Addict Behav. 2019;97:111–21. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5569740/
3. Olfson M, Blanco C, Wang S, Laje G, Correll CU. National trends in the mental health care of children, adolescents, and adults by office-based physicians. JAMA Netw Open. 2020;3(10):e2024676. Available from: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2772339
4. Read J, Cartwright C, Gibson K. Adverse emotional and interpersonal effects reported by 1829 New Zealanders while taking antidepressants. Psychiatry Res. 2014;216(1):67–73. Available from: https://www.tandfonline.com/doi/full/10.3109/13651500903447810
5. Read J, Williams J, Cohen D. Antidepressant withdrawal: a systematic review of the advice, experiences, and outcomes reported by users. Front Psychiatry. 2023;14:1087879. Available from: https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2023.1087879/full

Appendix
Model parameters

Random Forest (class weigh *balanced_subsample) – Minimising OOB :*

Paramters	**Grid**	**Tuned value – Male /Female**	**Tuned value – Female Anxious/ Non-Anxious**	**Tuned value – Male Anxious/ Non-Anxious**	**Tuned value – Biomarkers Male/Female**
Nb estimators	[500, 1000, 1250]	500/1250	1000/1250	1250	1250/1000
Criterion	“Gini”, “Entropy”	“Gini”	“Gini”	“Gini”	“Gini”
Max Depth	Range(3,8)	7	7	7	8
Min Sample Split	Range(2,5)	2	2	2	2
Min Sample Leaf	[5, 10, 15]	15/10	5	5	3
Max Features	“Sqrt”, “Log2”, “0.3”, “0.5”	“Sqrt”/“Log2”	“Log2”	“Log2”	“0.5”/”0.3”

Appendix
Model parameters:

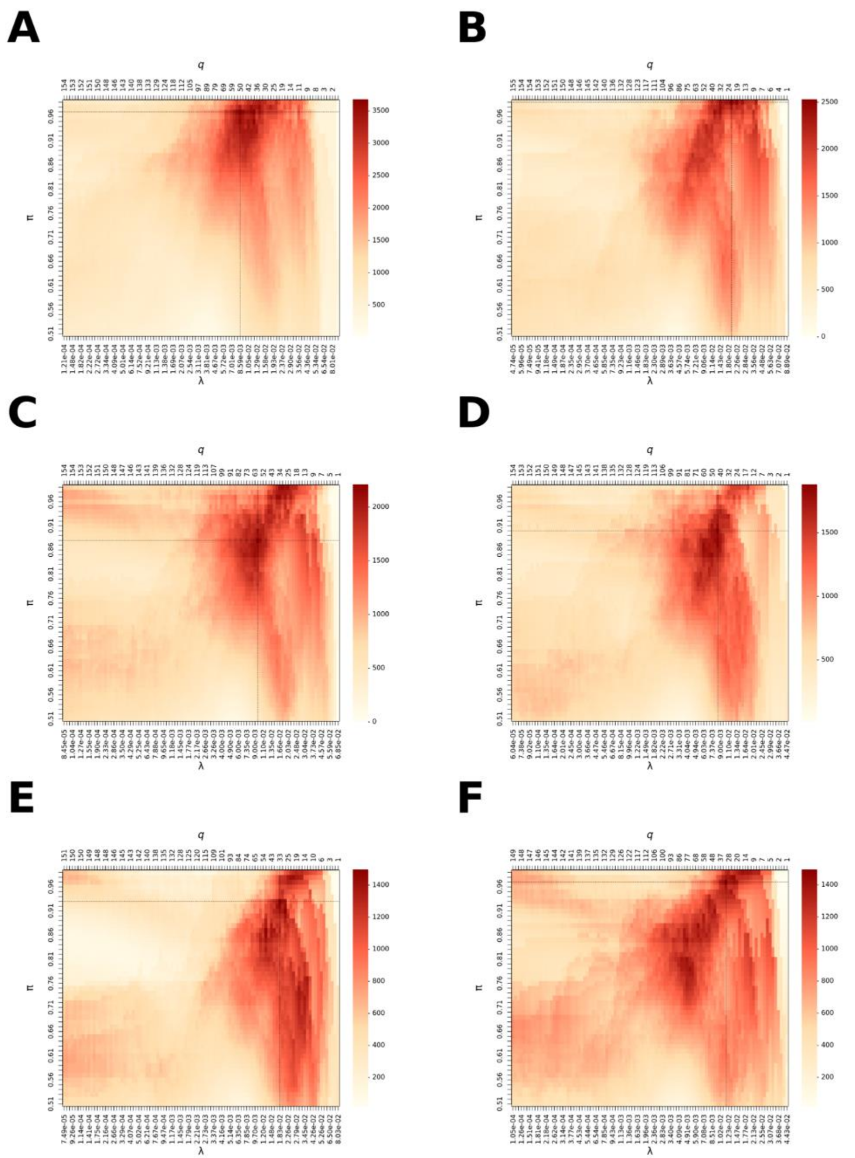
Xgboost – *10-fold cross validation maximisng AUC - objective is binary:logistic:*

Paramters	Grid	Tuned value – Male/ Female
Nb estimators	[100, 500, 1000]	500
Subsamples	[0.5, 0.7, 0.9]	0.5
Max Depth	[3,5,7,9]	3
Min Child Weight	[1, 3, 5]	5
Learning rate	[0.01 , 0.0575, 0.105 , 0.1525, 0.2]	0.01
Gamma	[0, 0.1, 0.2, 0.3]	0.2
Colsample by Tree	[0.5, 0.7, 0.9]	0.7

Appendix

Model Parameters – Stability Selection Lasso

Model	Lambda	Pi
Female (General, A)	0.009	0.97
Male (General, B)	0.019	0.99
Female (Anxiety, C)	0.010	0.88
Male (Anxiety, E)	0.018	0.93
Female (No Anxiety, D)	0.008	0.90
Male (No Anxiety, F)	0.012	0.97

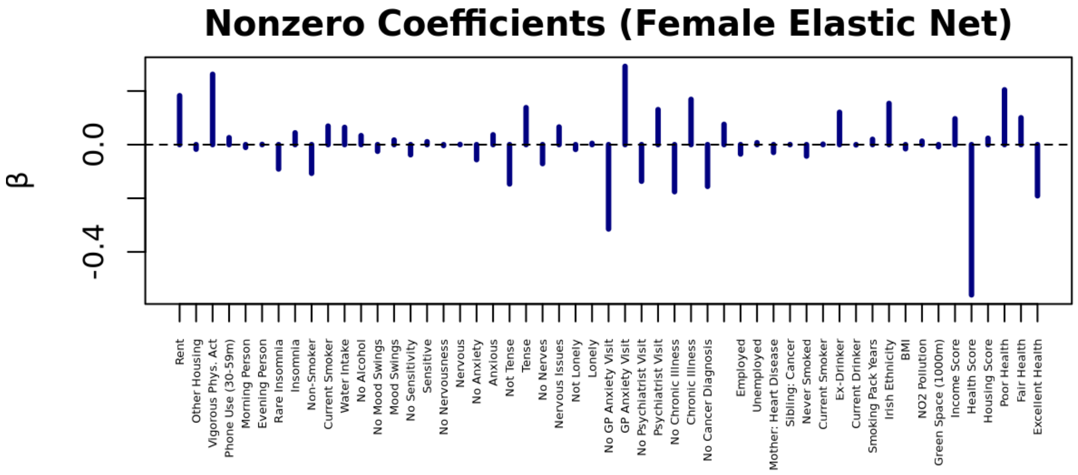


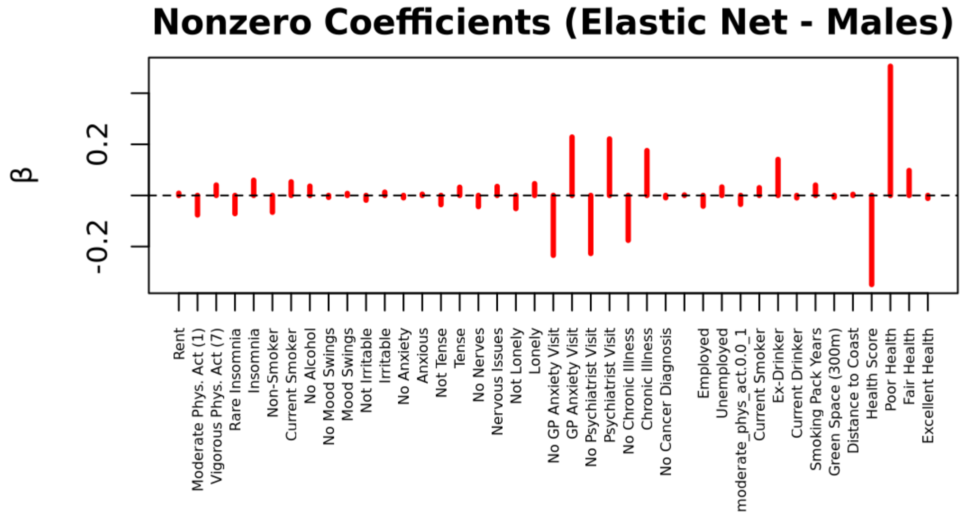
Appendix

Model Parameters – Elastic Net

The parameters are tuned using 10-fold cross validation, on AUC

Model	Lambda	Alpha
Female	0.012	0.326
Male	0.017	0.618





Appendix
Logistic Regression on Stably Selected Variables

Table 1 Results from logistic regression investigating the effect of stably selected features on long-term BZD prescription in anxious males, including Odds Ratio (95% Confidence Intervals) and associated p-values.

Variable	Odds Ratio	95% CI	p-value
(Intercept)	0.746	0.478, 1.160	0.194
Suffer from nerves (Yes)	1.270	0.916, 1.760	0.150
Loneliness/isolation (No)	0.842	0.584, 1.220	0.357
Long-standing illness/disability/infirmity (No)	0.620	0.439, 0.876	0.007
Current employment status (Employed)	0.808	0.580, 1.130	0.208
Health score	0.329	0.155, 0.590	0.001
Overall health rating: Poor	1.830	1.090, 3.110	0.022

Table 2 Results from logistic regression investigating the effect of stably selected features on long-term BZD prescription in non-anxious males, including Odds Ratio (95% Confidence Intervals) and associated p-values.

Variable	Odds Ratio	95% CI	p-value
(Intercept)	0.257	0.190 – 0.341	0.000
Long-standing illness/disability/infirmity (No)	0.628	0.467 – 0.848	0.002
Pack years of smoking	1.060	0.927 – 1.220	0.368
Health score	0.373	0.181 – 0.644	0.003
Overall health rating: Poor	1.250	0.653 – 2.310	0.480
Overall health rating: Fair	1.120	0.796 – 1.560	0.507

Table 3 Results from logistic regression investigating the effect of stably selected features on long-term BZD prescription in anxious females, including Odds Ratio (95% Confidence Intervals) and associated p-values.

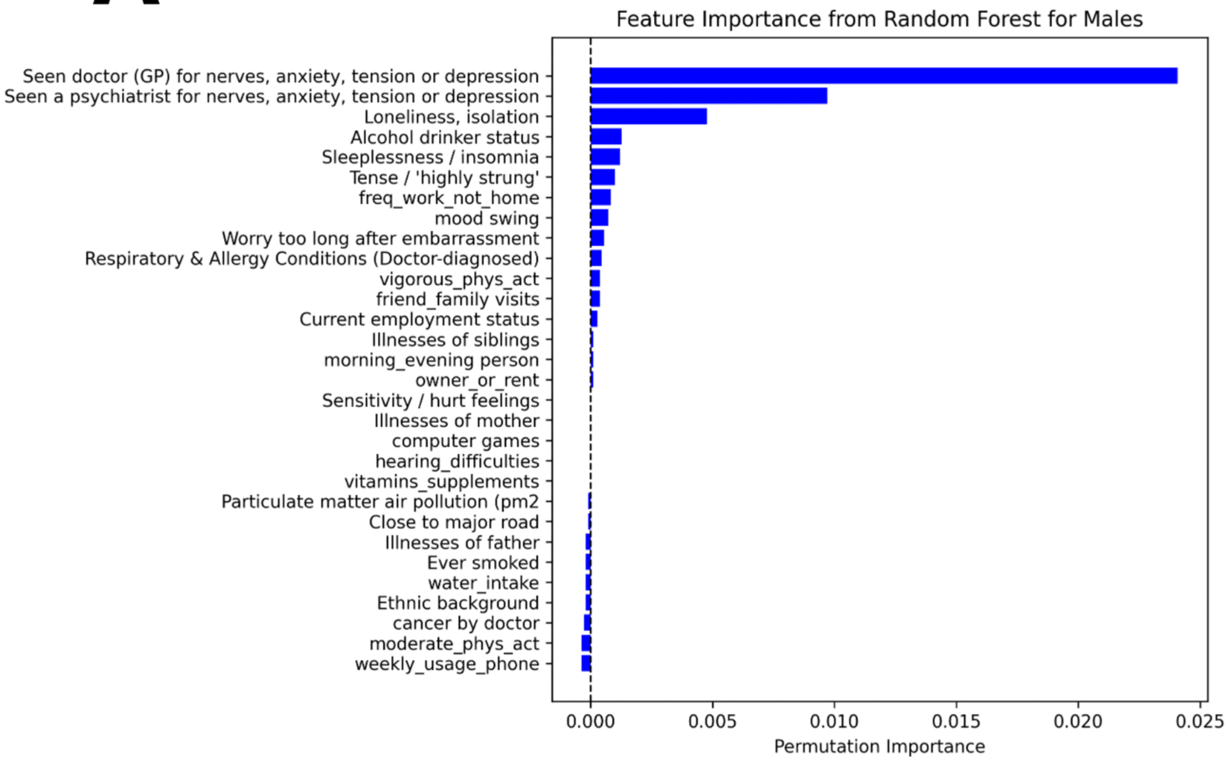
Variable	Odds Ratio	95% CI	p-value
(Intercept)	0.260	0.171, 0.392	0.000
Weekly phone usage: 30–59 mins	0.985	0.728, 1.330	0.922
Morning person	0.964	0.772, 1.200	0.746
Insomnia (Yes)	1.200	0.870, 1.660	0.274
Water intake	1.040	0.937, 1.160	0.433
Hurt feelings (No)	1.150	0.886, 1.480	0.298
Tense/highly strung (Yes)	1.330	1.030, 1.730	0.031
Suffer from nerves (Yes)	1.640	1.280, 2.120	0.000
Long-standing illness/disability (Yes)	1.520	1.210, 1.900	0.000
Cancer diagnosed (Yes)	1.380	0.946, 2.010	0.093
Employment status: Employed	0.976	0.782, 1.220	0.834
Income score	1.410	1.230, 1.620	0.000
Health score	0.248	0.143, 0.394	0.000
Overall health rating: Excellent	0.728	0.492, 1.060	0.106

Table 4 Results from logistic regression investigating the effect of stably selected features on long-term BZD prescription in non-anxious females, including Odds Ratio (95% Confidence Intervals) and associated p-values.

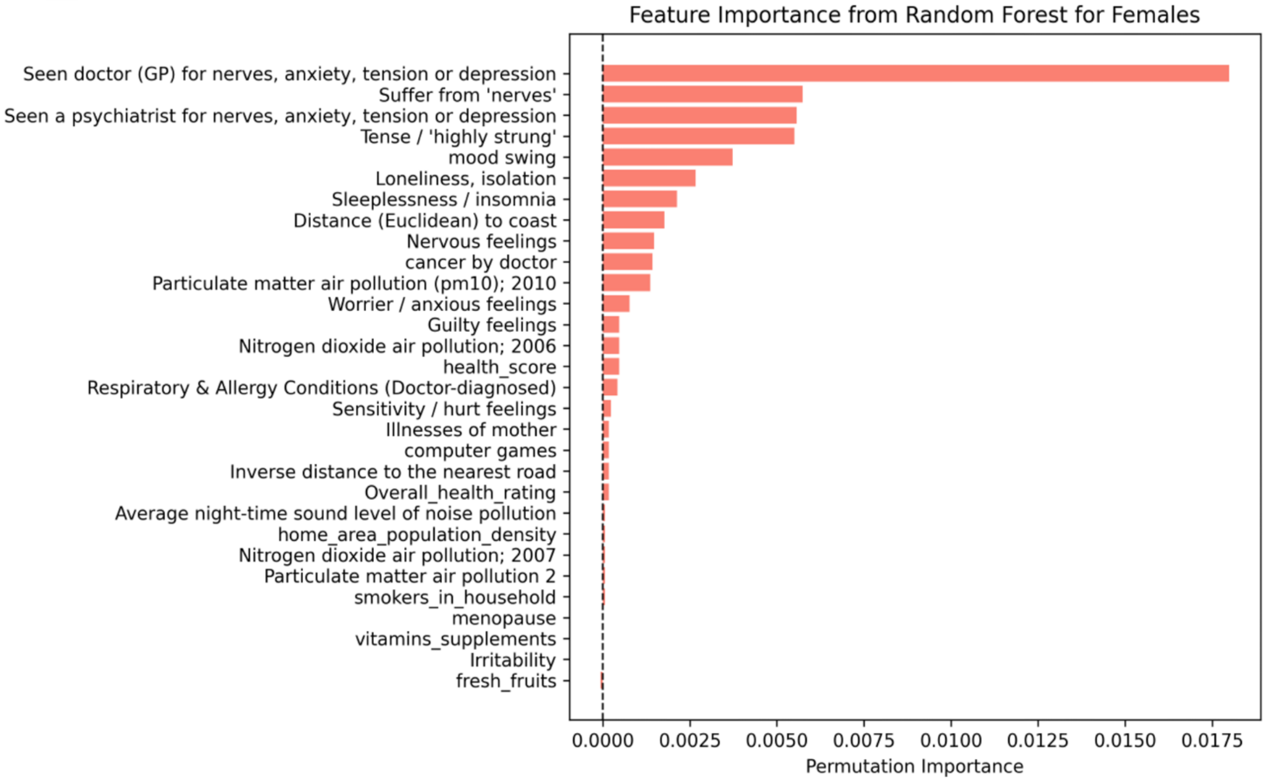
Variable	Odds Ratio	95% CI	p-value
(Intercept)	0.162	0.108, 0.237	0.000
Rents home	1.080	0.563, 1.970	0.803
Water intake	1.030	0.896, 1.170	0.695
Not tense/highly strung	0.673	0.464, 0.994	0.041
Long-standing illness/disability (Yes)	2.350	1.760, 3.120	0.000
Cancer diagnosed (Yes)	1.460	0.951, 2.180	0.075
Income score	1.130	0.940, 1.360	0.175
Health score	0.353	0.175, 0.603	0.001
Housing score	1.050	0.919, 1.200	0.455
Overall health rating: Fair	1.160	0.813, 1.630	0.406

Appendix
Permutation importance plot – Random Forest

A

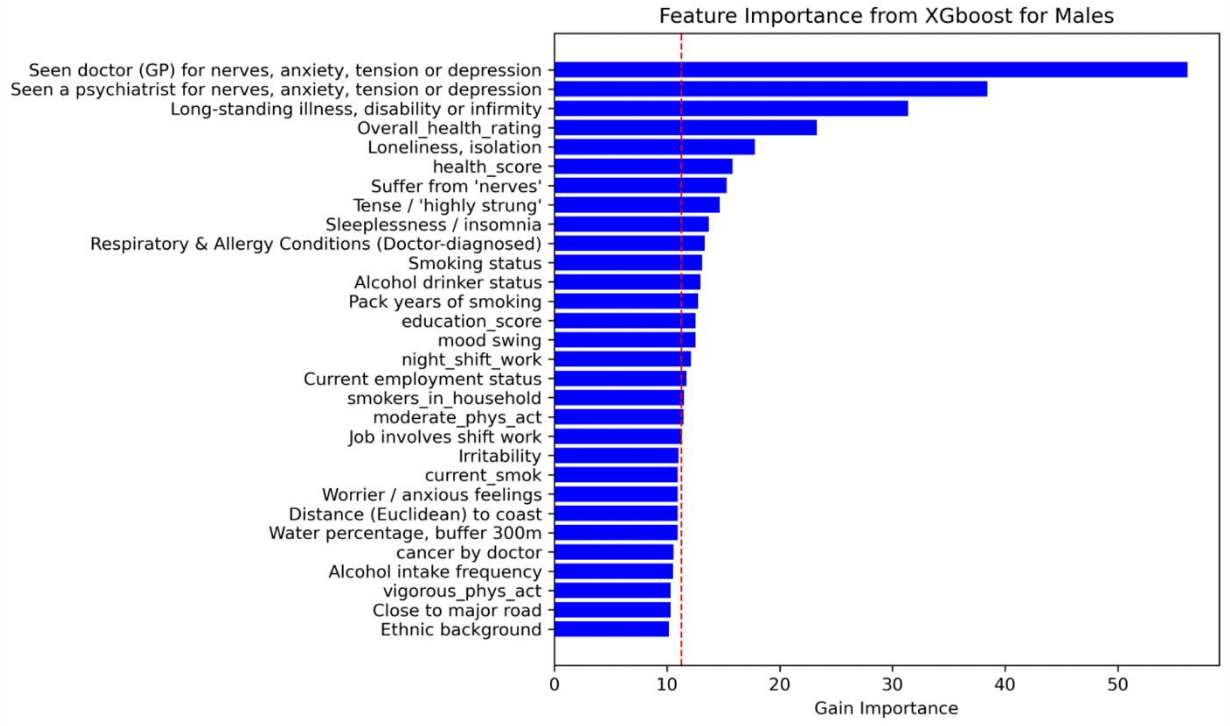


B



Appendix
Gain value plot – XGBoost

C



D