medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Psychiatric comorbid disorders of cognition: A machine learning approach using 1,159 UK Biobank participants

Chenlu Li, Delia A. Gheorghe, John E. Gallacher, Sarah Bauermeister

Department of Psychiatry and on behalf of Dementias Platform UK, Warneford Hospital, University of Oxford, Oxford, OX3 7JX

Corresponding author: Sarah Bauermeister sarah.bauermeister@psych.ox.ac.uk

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Background

Conceptualising comorbidity is complex and the term is used variously. Here, it is the coexistence of two or more diagnoses which might be defined as 'chronic' and, although they may be pathologically related, they may also act independently [1]. Of interest here is the comorbidity of common psychiatric disorders and impaired cognition.

Objectives

To examine whether anxiety and/or depression are important longitudinal predictors of impaired cognition.

Methods

UK Biobank participants used at three timepoints (n= 502,664): baseline, 1st follow-up (n= 20,257) and 1st imaging study (n=40,199). Participants with no missing data were 1,159 participants aged 40 to 70 years, 41% female. Machine learning (ML) was applied and the main outcome measure of reaction time intraindividual variability (cognition) was used.

Findings

Using the area under the Receiver Operating Characteristic (ROC) curve, the anxiety model achieves the best performance with an Area Under the Curve (AUC) of 0.68, followed by the depression model with an AUC of 0.64. The cardiovascular and diabetes model, and the demographics model have weaker performance in predicting cognition, with an AUC of 0.60 and 0.57, respectively.

Conclusions

Outcomes suggest psychiatric disorders are more important comorbidities of long-term cognitive change than diabetes and cardiovascular disease, and demographic factors. Findings suggest that

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

mental health disorders (anxiety and depression) may have a deleterious effect on long-term cognition and should be considered as an important comorbid disorder of cognitive decline.

Clinical implications

Important predictive effects of poor mental health on longitudinal cognitive decline should be considered in primary care.

Summary Box

What is already known about this subject? 3-4 bullet points

- Poor mental health is associated with cognitive deficits.
- One in four older adults experience a decline in affective state with increasing age.
- Conventional statistical approaches have inherent limitations as they are hypothesis-driven

What are the new findings? 3-4 bullet points

- Psychiatric disorders are important comorbid disorders of long-term cognitive change.
- Machine-learning methods such as sequence learning based methods are able to offer non-parametric joint modelling, allow for multiplicity of factors and provide prediction models that are more robust and accurate for longitudinal data
- The outcome of the RNN analysis found that anxiety and depression were stronger predictors of change IIV over time than either cardiovascular disease and diabetes or the covariate variables.

How might it impact on clinical practice in the foreseeable future?

The important predictive effect of mental health on longitudinal cognition should be noted and, its comorbidity relationship with other conditions such as cardiovascular disease likewise to be considered in primary care and other clinical settings

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Background

Conceptualising comorbidity is complex and the term is used variously. Here, we refer to it as the coexistence of two or more diagnoses which might be defined as 'chronic' and, although they may be pathologically related to each other, they may also act independently [1]. Of interest here is the comorbidity of common psychiatric disorders and impaired cognition.

Poor mental health is associated with cognitive deficits [2], with higher baseline levels of depressive symptoms predicting a steeper decline in delayed recall and global cognition at follow-up [3], and longitudinal slowing in processing speed [4]. Furthermore, at least one in four older adults experience a gradual decline in affective state [5] with increasing age, suggesting that along with age-related cognitive decline, there could be a comorbid relationship of unclear temporality [6]. Here we use a machine learning (ML) approach to investigate whether the presence of anxiety and depressive symptoms leads to greater cognitive decline.

Conventional statistical approaches have inherent limitations. As hypothesis-driven models, they are frequently pragmatically constrained by assumptions on data distributions and the linearity of associations, that are unlikely to be fully met, leading to a loss of statistical precision [7-11]. Also, high-dimensional data with noisy predictive variables, common in epidemiologic data are likely to possess slow convergence of estimators to the true value of a smooth function defined on a space of high dimension [12]. Therefore, preselection of variables is often necessary to avoid loss of important information which may be relevant for outcome prediction [11]. Although machine learning (ML) approaches are not without their limitations in terms of identifying mechanisms, they have advantages in identifying patterns of information useful for the prediction of an outcome, even when complex high-dimensional interactions exist [13]. Considering this, we apply ML techniques to

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

data from UK Biobank to examine whether anxiety and depression are important longitudinal predictors of cognitive decline.

Objective

To examine whether anxiety and/or depression are important longitudinal predictors of impaired cognition.

Methods

Governance

Ethical approval was granted to Biobank from the Research Ethics Committee - REC reference 11/NW/0382 [14]. All analyses were conducted using data from UK Biobank application 15008.

Population sample

UK Biobank recruited 502,664 participants who were registered with the UK National Health Service (NHS), aged between 37 and 73 years. For this analysis, participants with mental health and cognitive data at three time periods (recruitment, 5 year follow-up, and the 1st imaging sub-study) were used. For the purposes of this analysis the imaging sub-study data represents roughly a 10 year follow-up period from recruitment. All analyses were conducted listwise and the number of participants with no missing data across all measurement variables of interest were n = 1,159.

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Assessment

Details of the UK Biobank assessment procedures may be found elsewhere [14]. Cognitive performance was assessed using a 'stop-go' task where reaction time intraindividual variability (IIV) was used as a sensitive measure of cognitive change with greater IIV over time indicating decline in performance [15]. Mental health measures were four items from the PHQ9 [16] to assess depressive symptoms and the 12 item EPQ-R Neuroticism scale [17] as a proxy indicator for anxiety [18]. Other measures relevant to this analysis included self-report cardiovascular disease (CVD: including heart attack, angina and stroke) and diabetes, and observed body mass index (BMI) as indicators of wider co-morbidity [19 20]. Self-report smoking, alcohol consumption, and fruit and vegetable consumption were used as indicators of lifestyle. Self-report education, employment, ethnicity, and household income were used as socioeconomic indicators.

Analytic strategy

Data were modelled as follows. Both age and reaction time IIV were modelled as continuous variables. The reaction time IIV was calculated as the standard deviation of each participant's RTs over the test trials [21]. Participants with only one valid score at baseline or follow-up were omitted. The IIVs showed log–normal distributions and were natural log transformed. Mood and anxiety items were summed to obtain total scores for each and modelled as continuous variables. CVD and diabetes were modelled as binary indicators (present/absent). BMI was treated as a 3 level factor (normal (<25), overweight (25-29.9), obese (>29.9)). Lifestyle data were largely considered as binary indicators (smoker/non-smoker, <5 portions of fruit and vegetable per day / 5 or more portions per day), however alcohol was treated as a continuous variable (g/day). Socioeconomic data were considered as binary indicators (employed/not employed, college degree/no degree, white ethnicity/other) excepting income which was modelled as a 4 level factor (<£18,000, £18,000 - £30,999, £31,000 - £51,999; £52,000 - £100,000, >£100,000).

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Machine learning model

A recurrent neural network (RNN) is a class of artificial neural networks that is often utilized as an effective prediction tool for longitudinal biomedical data. By exhibiting temporal dynamic behaviour for time sequences, it allows for temporal dependencies between measurements. As a sequence learning based method, it is able to offer non-parametric joint modelling of longitudinal data [22 23].

Long short-term memory (LSTM) is an RNN architecture developed to deal with the exploding and vanishing gradient problem during backpropagation through time. It is able to effectively capture long-term temporal dependencies by employing a constant error carousel (memory cell) which maintains its state over arbitrary time intervals, and three nonlinear gating units (input, forget and output) which regulate the information flow into and out of the cell.[22 24-27]. This architecture is applied to longitudinally model the prediction abilities of psychiatric comorbidity disorders on cognition via sequence-to-sequence learning.

The dataset was partitioned into two non-overlapping subsets: 75% of the within-class subjects for training and 25% for testing. Data was rescaled [0 and 1] to meet the default hyperbolic tangent activation function of the LSTM model. The Adam version of stochastic gradient descent algorithm was applied to tune the weights of the network. This optimisation algorithm combines the advantages of Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp), thus being effective in handling sparse gradients on non-stationary problems (noisy). Furthermore, it compares favourably to other stochastic optimization methods in practice [28]. The model was set to fit 500 training epochs with a batch size set to the number of available training subjects. Mean absolute error (MAE) and Area under the Receiver Operating Characteristic (ROC) Curve (AUC) were used to assess model performance, respectively.

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Psychiatric comorbid disorders of cognition

To evaluate whether depression and anxiety are important comorbid disorders of cognition and, to compare the prediction abilities of diabetes and cardiovascular disease the model was run systematically to include:

I. Covariates only as reference model
II. Neuroticism scale + reference model
III. Depression scale + reference model
IV. Diabetes and cardiovascular disease + reference model

Given that major depressive disorder and anxiety have previously been considered to be comorbid disorders of impaired cognition, we expected predictions based on comorbid mental health disorders to outperform predictions based only on covariate variables.

A complete data analysis approach was used and all analyses were conducted on the Dementias Platform UK Data Portal [29] in Python 3.6.7 and MATLAB R2019.

Findings

Complete data were available for 1,159 Participants (Table 1). The sample was aged between 40 and 70 years at baseline, 41% female, and relatively healthy with few participants showing comorbidities and few reporting unhealthy behaviour such as smoking (5%).

The outcome of the RNN analysis found that anxiety and depression were stronger predictors of change IIV over time than either cardiovascular disease and diabetes or the covariate variables, as measured by the ROC curve (Figure 1 and Table 2). The AUC curve was 0.68 for anxiety, followed by the depression model with an AUC of 0.64. The cardiovascular and diabetes model and the demographics model had weaker performance with an AUC of 0.60 and 0.57, respectively. The

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

anxiety model prospectively identified 62.89% of patients who eventually reached high IIV (i.e., sensitivity), and 53.27% of patients who maintained low IIV (i.e., specificity). Correspondingly, the anxiety model had a positive predictive value (PPV) of 52.69%, and a negative predictive value (NPV) of 63.30%. The depression model achieved slightly smaller yet still significant predictive performance in comparison with the other models. The RNNs results suggest that the anxiety and depression model outperformed the reference model (Model I – covariates only). In general, the neuroticism and depression scales helped improve the accuracy in capturing the pattern of reaction time, and therefore may be more important predictors of impaired cognition over cardiovascular disease and diabetes. The results also suggest increased risk of multiple mental health disorders occurring together, suggesting depression and anxiety are potential comorbid disorders of cognition.

Discussion

A longitudinal analysis of the comorbid predictors of mental health disorders on cognitive decline was conducted using ML in a relatively healthy population sample. Evidence was found suggesting that anxiety and depression are important comorbid disorders of cognition but when cardiovascular disease, diabetes and other covariate predictors were included in the model these were not as important on long-term effect. This suggests that poor mental health has a significant deleterious effect on long-term cognition, and may be considered an important comorbid disorder of cognitive decline. The contribution of the present investigation is thereby three-fold.

First, this study used the community-based sample from UK Biobank with a general population large enough for identification of valid effects which might be implied across clinical and research settings. Clinical-based samples have been criticised for inconsistency in diagnoses when patients are treated over time in a variety of facilities or at different times in the same facility, or for failure to report all

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the comorbid diagnoses of a patient [30]. In comparison, community-based samples are based on a system in which a practical number of criteria are rated, or collected and assessed on every patient, providing unbiased prevalence or incidence rates of comorbidity, or unbiased estimates of risk factors for comorbidity [30].

Second, this study investigates the psychiatric comorbid disorders of cognition in a comprehensive ML design which was applied to examine if psychiatric disorders as measured by self-report scales are important predictors of impaired cognition. Unlike conventional statistical techniques which are hypothesis-driven, the sequence learning based method is able to offer non-parametric joint modelling, allow for multiplicity of factors and provide prediction models that are more robust and accurate for longitudinal data [22].

Third, this study is conducted longitudinally. Disease progression modelling using longitudinal data may provide clinicians with improved tools for diagnosis and monitoring of diseases. We apply the widely used LSTM network to capture the long-term temporal dependencies among measurements without making parametric assumptions about cognitive trajectories. To our knowledge there are no studies which have employed the ML techniques presented here to investigate the comorbidity of mental health disorders and longitudinal cognitive decline using cardiovascular disease and diabetes as comorbidity covariates. Utilising ML methodologies for clinical data such as psychiatric outcomes may provide additional information which is not possible with traditional statistical methodologies.

Clinical Implications

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The implication of this work is that the important predictive effect of mental health on longitudinal cognition should be noted and, its comorbidity relationship with other conditions such as cardiovascular disease likewise to be considered in primary care and other clinical settings.

Acknowledgements

This is a DPUK supported project with all analyses were conducted on the DPUK Data Portal, constituting part 1 of DPUK Application 0132.

The Medical Research Council supports DPUK through grant MR/L023784/2.

CL was funded by DPUK to complete this analysis.

Authors' Contributions

CL, JG and SB conceptualised the idea. CL analysed the data, CL, DG, JG and SB interpreted the data, CL wrote the 1st draft of the manuscript. DG and SB commented, edited and proofread the manuscript. All authors read and approved the final manuscript.

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

1. Kraemer HC. Statistical issues in assessing comorbidity. *Stat Med* 1995;14(8):721-33. [published Online First: 1995/04/30]
2. Bauermeister S, Bunce D. Poorer mental health is associated with cognitive deficits in old age. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn* 2015;22(1):95-105. doi: 10.1080/13825585.2014.893554 [published Online First: 2014/03/13]
3. Panza F, Frisardi V, Capurso C, et al. Late-life depression, mild cognitive impairment, and dementia: possible continuum? *Am J Geriatr Psychiatry* 2010;18(2):98-116. doi: 10.1097/JGP.0b013e3181b0fa13 [published Online First: 2010/01/28]
4. Bunce D, Batterham PJ, Christensen H, et al. Causal associations between depression symptoms and cognition in a community-based cohort of older adults. *Am J Geriatr Psychiatry* 2014;22(12):1583-91. doi: 10.1016/j.jagp.2014.01.004 [published Online First: 2014/02/08]
5. Arve S, Tilvis RS, Lehtonen A, et al. Coexistence of lowered mood and cognitive impairment of elderly people in five birth cohorts. *Aging (Milano)* 1999;11(2):90-5. [published Online First: 1999/07/01]
6. John A, Patel U, Rusted J, et al. Affective problems and decline in cognitive state in older adults: a systematic review and meta-analysis. *Psychol Med* 2019;49(3):353-65. doi: 10.1017/S0033291718001137 [published Online First: 2018/05/25]
7. Breiman L. Random forests. *Machine learning* 2001;45(1):5-32.
8. Breiman L. Classification and regression trees: Routledge 2017.
9. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning: Springer series in statistics New York 2001.
10. Ripley BD. Pattern recognition and neural networks: Cambridge university press 2007.
11. Shouval R, Bondi O, Mishan H, et al. Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. *Bone marrow transplantation* 2014;49(3):332.
12. Munshi UM, Chaudhuri BB. Multimedia information extraction and digital heritage preservation: World Scientific 2011.
13. Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry* 2016;3(3):243-50.
14. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* 2015;12(3):e1001779.
15. Bunce D, Bauermeister S. Intraindividual Reaction Time Variability, Attention, and Age-Related Outcomes. Oxford Research Encyclopedia of Psychology: Oxford University Press 2019.
16. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16(9):606-13. doi: 10.1046/j.1525-1497.2001.016009606.x [published Online First: 2001/09/15]
17. Eysenck SBG, Eysenck HJ, Barrett P. A Revised Version of the Psychoticism Scale. *Pers Indiv Differ* 1985;6(1):21-29. doi: Doi 10.1016/0191-8869(85)90026-1
18. Ormel J, Jeronimus BF, Kotov R, et al. Neuroticism and common mental disorders: meaning and utility of a complex relationship. *Clin Psychol Rev* 2013;33(5):686-97. doi: 10.1016/j.cpr.2013.04.003 [published Online First: 2013/05/25]
19. Hassing LB, Hofer SM, Nilsson SE, et al. Comorbid type 2 diabetes mellitus and hypertension exacerbates cognitive decline: evidence from a longitudinal study. *Age Ageing* 2004;33(4):355-61. doi: 10.1093/ageing/afh100 [published Online First: 2004/05/12]
20. Biessels GJ, Deary IJ, Ryan CM. Cognition and diabetes: a lifespan perspective. *Lancet Neurol* 2008;7(2):184-90. doi: 10.1016/S1474-4422(08)70021-8 [published Online First: 2008/01/22]
21. Dykiert D, Der G, Starr JM, et al. Age differences in intra-individual variability in simple and choice reaction time: systematic review and meta-analysis. *PLoS One* 2012;7(10):e45759.

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

22. Ghazi MM, Nielsen M, Pai A, et al. Robust training of recurrent neural networks to handle missing data for disease progression modeling. *arXiv preprint arXiv:180805500* 2018
23. Pearlmutter BA. Learning state space trajectories in recurrent neural networks. *Neural Computation* 1989;1(2):263-69.
24. Gers FA, Schmidhuber E. LSTM recurrent networks learn simple context-free and context-sensitive languages. *IEEE Transactions on Neural Networks* 2001;12(6):1333-40.
25. Gers FA, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM. 1999
26. Greff K, Srivastava RK, Koutník J, et al. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* 2017;28(10):2222-32.
27. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural computation* 1997;9(8):1735-80.
28. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980* 2014
29. Bauermeister S, Orton C, Thompson S, et al. The Dementias Platform UK (DPUK) Data Portal. *BioRxiv* Preprint doi: 10.1101/582155
30. Cloninger CR. Implications of comorbidity for the classification of mental disorders: the need for a psychobiology of coherence. *Psychiatric diagnosis and classification* 2002;79:106.

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 1 Descriptive statistics

	Baseline			First Assessment			Imaging		
	%	IIV mean	IIV SD	%	IIV mean	IIV SD	%	IIV mean	IIV SD
Panel A – Neuroticism EPQ-R Scale									
"Does your mood often go up and down?"									
Yes	0.33	72.23	53.45	0.30	71.46	47.22	0.28	79.44	54.80
No	0.67	72.22	47.46	0.70	73.30	48.61	0.72	74.40	50.80
"Do you ever feel just miserable for no reason?"									
Yes	0.33	68.95	46.58	0.30	70.77	41.79	0.27	78.51	51.95
No	0.67	73.86	50.84	0.70	73.57	50.62	0.73	74.80	51.99
"Are you an irritable person?"									
Yes	0.25	73.19	53.07	0.24	73.57	52.16	0.22	73.03	51.66
No	0.75	71.90	48.27	0.76	72.48	46.88	0.78	76.62	52.08
"Are your feelings easily hurt?"									
Yes	0.47	75.65	50.61	0.46	73.70	51.53	0.43	75.39	48.10
No	0.53	69.14	48.30	0.54	71.94	45.19	0.57	76.14	54.76
"Do you often feel fed-up?"									
Yes	0.28	70.34	47.89	0.27	69.78	46.52	0.23	79.35	53.21
No	0.72	72.95	50.11	0.73	73.86	48.77	0.77	74.75	51.59
"Would you call yourself a nervous person?"									
Yes	0.18	78.47	58.24	0.17	72.72	52.88	0.15	77.37	54.49
No	0.82	70.87	47.31	0.83	72.75	47.20	0.85	75.55	51.56
"Are you a worrier?"									
Yes	0.45	72.09	46.50	0.46	74.02	50.93	0.43	74.85	50.18
No	0.55	72.34	51.87	0.54	71.67	45.74	0.57	76.53	53.31
"Would you call yourself tense or highly strung?"									
Yes	0.10	74.06	47.79	0.10	79.92	61.00	0.08	75.54	49.11
No	0.90	72.01	49.70	0.90	71.97	46.55	0.92	75.84	52.26
"Do you worry too long after an embarrassing experience?"									
Yes	0.45	69.92	40.67	0.46	72.73	51.27	0.47	73.16	47.34
No	0.55	74.10	55.61	0.54	72.75	45.44	0.53	78.16	55.69
"Do you suffer from nerves?"									
Yes	0.16	75.07	53.72	0.14	70.94	50.81	0.12	68.17	44.55
No	0.84	71.69	48.67	0.86	73.04	47.76	0.88	76.83	52.82
"Do you often feel lonely?"									
Yes	0.12	72.40	53.25	0.12	70.47	36.80	0.11	75.49	50.79
No	0.88	72.20	48.99	0.88	73.05	49.55	0.89	75.86	52.15
"Are you often troubled by feelings of guilt?"									
Yes	0.23	69.12	37.08	0.22	73.61	52.26	0.24	76.88	51.02
No	0.77	73.17	52.67	0.78	72.49	46.96	0.76	75.49	52.30
Panel B – PHQ-4 Depression Scale									

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

"frequency of depressed mood"									
Not at all	0.83	72.35	51.21	0.86	73.84	49.50	0.87	76.02	52.46
Several days	0.15	72.11	40.91	0.12	65.65	39.67	0.11	73.25	47.57
More than	0.01	72.45	43.13	0.01	76.43	38.61	0.01	89.04	60.27
Nearly	0.01	60.70	24.81	0.01	60.88	28.68	0.00	54.53	36.12
"frequency of unenthusiasm / disinterest"									
Not at all	0.87	71.87	50.42	0.89	73.33	49.04	0.90	76.05	52.30
Several days	0.11	74.18	42.86	0.09	70.66	42.61	0.08	70.89	48.15
More than	0.01	65.79	35.19	0.01	55.48	25.83	0.01	89.56	59.17
Nearly	0.01	91.88	53.72	0.01	58.10	34.68	0.00	101.0	40.76
"frequency of tenseness / restlessness"									
Not at all	0.77	72.24	51.08	0.81	74.06	50.03	0.84	76.04	51.73
Several days	0.20	73.54	45.06	0.16	67.73	39.39	0.15	73.58	54.13
More than	0.02	64.56	31.23	0.01	64.34	33.98	0.01	92.90	48.79
Nearly	0.01	55.74	25.44	0.01	51.91	25.44	0.01	73.27	36.85
"frequency of tiredness / lethargy"									
Not at all	0.57	73.37	51.06	0.62	72.33	47.73	0.66	75.98	52.21
Several days	0.37	70.00	47.68	0.34	73.90	50.51	0.30	74.74	50.97
More than	0.04	78.10	49.18	0.03	72.75	39.61	0.03	82.03	62.78
Nearly	0.02	69.37	38.64	0.02	64.72	25.90	0.02	77.33	42.54
Panel C – Other Comorbid Disorders									
Heart attack									
Yes	0.01	62.04	33.94	0.01	75.11	55.66	0.02	65.52	51.29
No	0.99	72.33	49.63	0.99	72.71	48.09	0.98	76.02	52.00
Angina									
Yes	0.01	60.13	24.80	0.01	72.81	44.78	0.01	64.62	27.90
No	0.99	72.35	49.68	0.99	72.74	48.24	0.99	75.94	52.19
Stroke									
Yes	0.01	79.49	38.78	0.01	89.73	51.71	0.01	71.77	45.78
No	0.99	72.17	49.58	0.99	72.55	48.13	0.99	75.86	52.06
High blood pressure									
Yes	0.18	73.33	43.98	0.22	73.66	53.68	0.22	77.31	48.90
No	0.82	71.98	50.63	0.78	72.48	46.54	0.78	75.39	52.85
Diabetes									
Yes	0.02	64.91	23.10	0.03	69.54	43.77	0.03	99.29	51.22
No	0.98	72.38	49.90	0.97	72.83	48.30	0.97	75.11	51.86
Panel D - Covariates									
Gender									
Male	0.59	72.31	48.22	0.59	71.37	48.16	0.59	75.38	54.59
Female	0.41	72.11	51.31	0.41	74.71	48.19	0.41	76.45	48.05
Education									
College	0.56	69.28	44.87	0.56	72.94	49.01	0.58	73.09	52.04

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

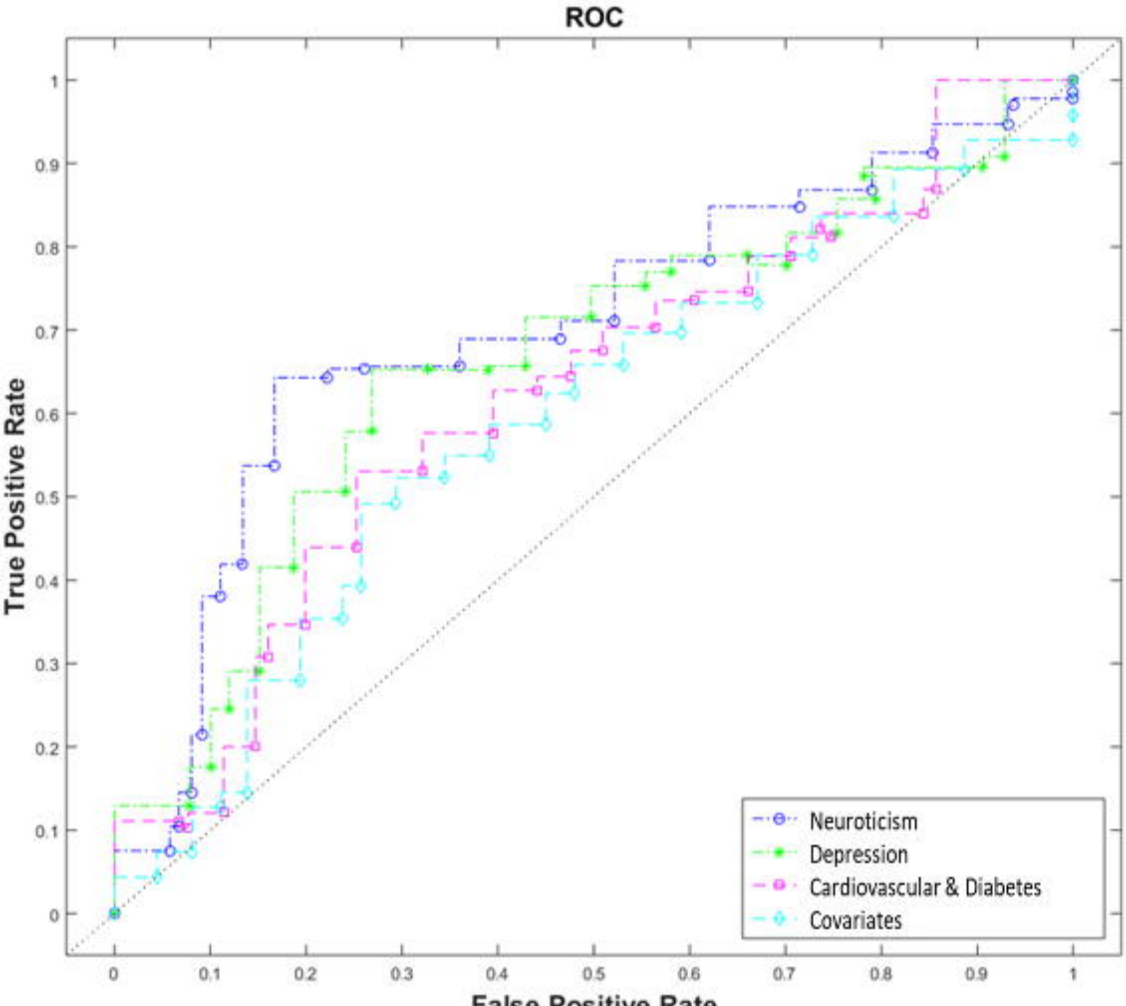
Other	0.44	75.91	54.56	0.44	72.50	47.16	0.42	79.61	51.72
BMI									
Normal	0.41	72.96	54.50	0.42	74.38	49.24	0.45	77.38	55.56
Overweight	0.43	73.15	48.36	0.42	73.50	48.84	0.40	74.47	49.82
Obese	0.15	67.63	36.99	0.15	66.13	42.80	0.15	74.70	46.21
Household Income									
<£18,000	0.10	84.88	74.74	0.12	77.19	42.61	0.11	84.06	58.16
(£18,000,£30,99	0.22	76.05	44.03	0.27	74.34	47.62	0.27	78.97	47.54
(£31,000,£51,99	0.30	72.51	47.30	0.30	73.71	50.86	0.30	69.63	45.74
(£52,000,£100,0	0.28	68.74	47.82	0.23	73.64	52.30	0.23	79.28	60.62
>£100,000	0.09	59.26	35.36	0.08	55.67	29.28	0.09	68.53	50.14
Smoking status									
No	0.95	63.09	34.48	0.96	72.57	48.36	0.97	75.62	51.64
Yes	0.05	72.73	50.16	0.04	77.31	43.30	0.03	82.41	62.88
Employment status									
Employed	0.67	68.10	45.55	0.51	68.90	46.22	0.42	73.22	49.83
Other	0.33	80.64	55.83	0.49	76.80	49.88	0.58	77.73	53.48
Ethnicity									
White	0.98	72.28	49.41	0.98	72.74	47.75	0.98	75.42	51.60
Other	0.02	69.32	55.05	0.02	72.66	68.22	0.02	94.84	66.43
Vegetable and fruit intake									
≥ 5 unit per	0.76	72.27	49.40	0.74	72.98	47.49	0.79	74.99	49.69
< 5 unit per	0.24	72.07	49.88	0.26	72.06	50.13	0.21	78.87	59.67
					Mean				
Age		55.81			60.09			62.62	
Alcohol Consumption (g/day)		18.39			16.43			16.39	

medRxiv preprint doi: https://doi.org/10.1101/19006189; this version posted March 2, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 2 Model performance for comorbid longitudinal predictors of IIV

	Anxiety	Depression	Cardiovascular & Diabetes	Covariates
AUC	0.68	0.64	0.60	0.57
Sensitivity	62.89%	62.72%	54.91%	53.56%
Specificity	53.27%	52.82%	55.34%	53.80%
PPV	52.69%	50.96%	49.31%	47.10%
NPV	63.30%	62.45%	59.82%	59.50%

AUC = area under receiver characteristics curve. PPV = positive predictive value. NPV = negative predictive value.


ROC
True Positive Rate
False Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Neuroticism
Depression
Cardiovascular & Diabetes
Covariates