Assessing psychiatric comorbid disorders of cognition:

A machine learning approach using UK Biobank data

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

Abstract

Background

Using UK Biobank data from baseline to the imaging study, our aim was to investigate whether participants with comorbid psychiatric disorders suffer from higher risk of impaired cognition over time compared to those without. There are several conventional statistical techniques to examine whether this pattern indicates the increased risk of comorbidity of the aforementioned, however, they carry certain inherent limitations. Machine learning (ML) has shown specific advantages in examining potential predictors simultaneously in an unbiased manner, especially its ability to identify patterns of information within useful features for the prediction of an outcome of interest. We applied ML techniques on UK Biobank data to examine whether anxiety and/or depression are important longitudinal predictors of impaired cognition.

Methods

We used data from UK Biobank (n = 1.158) across three time waves to longitudinally assess the effects of comorbidity (anxiety and depression, measured through self-report scales) on cognition. Theta (Θ) measures for each mental health scale were computed using item response theory (IRT) and intraindividual variability of reaction time performance (IIV) - raw standard deviation - was used as a measure of cognitive performance. First, comorbidity information was summarized to show the variation of impaired cognition for participants with and without psychiatric disorders. A machine learning (ML) approach was then applied to examine if psychiatric disorders and other putative covariate markers may be important predictors of long-term cognitive decline.

Results

For participants with anxiety, the percentage of having impaired cognition constantly increases through time from 35.09% to 42.15%; in contrast, an increase with less magnitude

is found for participants without anxiety from 36.10% to 40.03%. Likewise, for participants with depression, the percentage of having impaired cognition constantly increases through time from 35.49% to 41.67%; in contrast, an increase with less magnitude is found for participants without depression from 35.64% to 40.52%. Using the area under the Receiver Operating Characteristic (ROC) curve it was observed that the anxiety model achieved the best performance among all models, 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 had relatively weak performance in predicting cognition, with an AUC of 0.60 and 0.57, respectively.

Conclusions

Using data from UK Biobank, this study provides empirical evidence which suggests that psychiatric disorders are important comorbidities of cognitive decline. Furthermore, when other comorbidities were included in the model these were not as important on long-term effect. When the recurrent neural networks were trained for the psychiatric disorder features (anxiety and depression) they showed improved performance in predicting cognition in comparison with cardiovascular disease, diabetes and demographic factors. These findings suggest that mental health disorders (anxiety and depression) have a deleterious effect on long-term cognition, and may be considered an important comorbid disorder of cognitive decline. The implications of this work is that the important predictive effect of poor mental health on longitudinal cognitive decline should be considered in both research and clinical settings.

Introduction

The concept of comorbidity was initially operationalised by Feinstein when he was concerned with the medicating confounds of treating patients with diseases such as rheumatic fever who simultaneously suffered from multiple diseases (Feinstein, 1970). Conceptualising comorbidity, however, has not proved so straightforward and since Feinstein, the absence of a true definition has led to it constituting a heterogeneous term in the absence of its own nosology. Moreover, it is used synonymously with multimorbidity, although the latter is more commonly referred to when the co-existing disorders (diseases) are non-defined in their own right, with comorbidity in contrast, seen as a 'coexistence' of diseases (Meghani et al., 2013). Moreover, in psychiatry, psychology, and mental health counselling, comorbidity has been referred to as the presence of more than one diagnosis occurring in an individual at the same time (Cloninger, 2002). Here, we reflect on the definition and refer to it as the coexistence of not only two or more diseases but also, diseases which might be defined as 'chronic' and although they might be pathologically related to each other, they also act independently (Kraemer, 1995).

Our interest is in the comorbidity of psychiatric disorders and impaired cognition. Research suggests that poorer mental health is associated with cognitive deficits (Bauermeister & Bunce, 2015), with higher baseline levels of depressive symptoms predicting a steep decline in delayed recall and global cognition at follow-up (Panza et al., 2010) and longitudinal slowing in processing speed (Bunce, Batterham, Christensen, & Mackinnon, 2014). Furthermore, with at least one in four older adults experiencing a gradual decline in affective state with increasing age, there is a suggestion that along with age-related cognitive decline, there could be a comorbid relationship of unclear temporality (Arve, Tilvis, Lehtonen, Valvanne, & Sairanen, 1999). Using UK Biobank data from baseline to the imaging study, our aim was to investigate whether those participants with anxiety and/or depression suffer from higher risk of impaired cognition over time compared to those without. There are several

conventional statistical techniques to examine whether this pattern indicates the increased risk of comorbidity of the aforementioned, however, they carry certain inherent limitations. First, as hypothesis-driven models, they tend to have assumptions on data distributions. Above all, the assumption of normally distributed independence of attributes or linear data behaviour are often not met in large registries incorporating neuropsychological data. With the presence of nonlinear high-order interaction effects, the statistical inference might lose precision (Breiman, 2001, 2017; Friedman, Hastie, & Tibshirani, 2001; Ripley, 2007; Shouval et al., 2014). Second, high-dimensional data with noisy predictive variables, common in such data, is likely to possess slow convergence of estimators to the true value of a smooth function defined on a space of high dimension (Munshi & Chaudhuri, 2011). Therefore, preselection of variables is often necessary to avoid loss of important information which may be relevant for outcome prediction (Shouval et al., 2014).

On the other hand, machine learning (ML) has shown certain advantages in examining potential predictors simultaneously in an unbiased manner - especially its ability to identify patterns of information within useful features for the prediction of an outcome of interest, even when complex high-dimensional interactions exist (Chekroud et al., 2016). A recent report suggested that ML models developed from self-reports about characteristics of incident episodes and comorbidities among respondents with lifetime major depressive disorder (MDD) predicted MDD persistence, chronicity and severity with good accuracy (van Loo et al., 2014; Wardenaar et al., 2014). Although ML methods have successfully been used to develop risk-prediction schemes in other areas of medicine (Chang, Chen, Chang, Chung, & Lai, 2012; Chao et al., 2012), applications to psychiatric comorbid disorders have so far relied on small samples and thin predictor sets, failing to realize the full potential of the methods. Considering this, here we apply ML techniques on UK Biobank data to examine whether anxiety/depression are important longitudinal predictors of impaired cognition.

Methods

Dataset description

The UK Biobank recruited 502,664 participants who were registered with the UK National Health Service (NHS), aged between 37 and 73 years. Participants undertook comprehensive mental health, cognitive, lifestyle, biomedical, demographic and physical assessments at 22 assessment centres throughout the UK between 2006 and 2010. Follow-up assessments were conducted for approximately 20,000 participants in 2013 (Sudlow et al., 2015) and further follow-up assessments were collected for those participants taking part in the imaging study which is ongoing. Ethical approval was granted to Biobank from the Research Ethics Committee - REC reference 11/NW/0382 (Sudlow et al., 2015). All analyses were conducted on the Dementias Platform UK (DPUK) Data Portal (Bauermeister et al., Preprint) using data from UK Biobank application 15008. The analysis was conducted in Python 3.6.7 and MATLAB R2015b.

We utilise a dataset which includes three time waves in UK Biobank: baseline (time wave 1), first assessment (time wave 2) and imaging visit assessment data (time wave 3), to longitudinally examine the role of depression and anxiety as comorbid disorders of impaired cognition. Only participants who completed the three visits were included in the study to allow two visits for performing sequence learning and an additional visit for backpropagation (Ghazi et al., 2018). Participants with missing observations at any time wave were removed to obtain 1,158 subjects. Descriptive statistics are reported in Table 1.

Outcome measures

Cognitive performance was assessed using a ‘stop-go’ RT task in which participants were shown two cards simultaneously on a computer screen. Each card had a symbol on it and participants were asked to respond as quickly as possible using a button-box, if both symbols matched. Reaction time, from the presentation of the cards to the press of the button, was recorded in milliseconds (ms). Each participant was presented with 12 pairs of cards, the

first five of which were training sets. Data from these trials were discarded. Of the seven test trials, cards with matching symbols were presented on four occasions selected at random. Consistent with established methods (e.g., Lyall et al., 2016), participants with a reaction time less than 50ms indicating anticipation of the stimulus were discarded, and those with reaction times greater than 120ms were treated as the target stimulus had been withdrawn at this point (Piumatti, Moore, Berridge, Sarkar, & Gallacher, 2018). Intra-individual variability (IIV) was calculated as the standard deviation of each participant's RTs over the test trials (Dykiert, Der, Starr, & Deary, 2012). Participants with only one valid score at baseline or follow-up were omitted. IIV showed log–normal distributions and were natural log transformed.

Comorbid disorders

This paper focused on two common psychiatric comorbid disorders: MDD and anxiety. Participants completed a selection of mental health scales presented on a touchscreen computer, including the Patient Health Questionnaire four-item version (PHQ-4) to capture depression (Kroenke, Spitzer, Williams, & Lowe, 2009). Participants were required to rate on a four-point Likert scale from zero (not at all) to three (nearly every day), their response to four items: "frequency of depressed mood", "frequency of unenthusiasm / disinterest", "frequency of tenseness / restlessness" and "frequency of tiredness / lethargy". An anxiety proxy measure was utilised using the 12-item neuroticism EPQ-R scale (Eysenck, Eysenck, & Barrett, 1985) which suggests emotional instability and reactiveness. Participants were required to answer, "yes", "no", "I don't know" or "I do not wish to answer" in response to the 12 questions: "Does your mood often go up and down?", "Do you ever feel just miserable for no reason?", "Are you an irritable person?", "Are your feelings easily hurt?", "Do you often feel fed-up?", "Would you call yourself a nervous person?", "Are you a worrier?", "Would you call yourself tense or highly strung?", "Do you worry too long after an embarrassing experience?", "Do you suffer from nerves?", "Do you often feel lonely?", and "Are you often troubled by feelings of guilt?".

Covariates

Diabetes and cardiovascular disease including heart attack, angina and stroke were included as binary variables (yes / no). They were included as proxies for other potential comorbid disorders (Biessels, Deary, & Ryan, 2008; Hassing et al., 2004). Other common confounders of cognition and psychiatric disorders were also included as covariates (Arve et al., 1999; Feart, Samieri, & Barberger-Gateau, 2010; Mathew, Hogarth, Leventhal, Cook, & Hitsman, 2017; Piumatti et al., 2018). Educational attainment was included as a binary variable (with or without a university or college degree), as was tobacco use status (smoker / non-smoker), ethnicity (white / other), employment status (employed / not), and consumption of fruit and vegetables (below 5 portions a day compared to equal or above). Body mass index (BMI) was included as two binary variables with normal weight defined as ≤ 24.9kg/m2 compared to overweight 25.0 to 29.9kg/m2 and obese ≥ 30kg/m2. Categorical information on annual household income was available: less than £18,000; £18,000 to £30,999; £31,000 to £51,999; £52,000 to £100,000 and greater than £100,000. Age was included as a continuous variable. Alcohol consumption was calculated as the summation of the multiplication of the average number of consumed alcoholic drinks by the average grams of alcohol contained in each type of drink using the UK Food Standard Agency's guidelines (Roe, Pinchen, Church, & Finglas, 2015).

Analytic Strategy

In this study, we examined whether participants have longitudinal increased risk of two mental health disorders (MDD and anxiety) occurring together, in two steps:

Step 1 – Summarise the comorbidity information

We present an overview of whether mental health disorders and impaired cognition are likely to occur together by examining how cognition varies for participants with/without disorders, over time, using two approaches.

First, we follow Elixhauser *et al*. (1998) to stratify participants into two groups: those with a disorder (depression/anxiety) and those without. The percentage of participants with high/low IIV (above/below average) is calculated for each stratified disorder group in each time wave. Second, we follow Charlson *et al*. (1987) to measure the unadjusted relative risks of comorbidity. This assesses the risk of patients' impaired cognition with a given comorbid condition (anxiety/depression in this case) regardless of the presence of other comorbid diseases. It is calculated as the division of the proportion of patients with anxiety/depression who have impaired cognition by the proportion of patients without anxiety/depression who have impaired cognition.

Theta (ϴ) measures were computed for depression and anxiety separately using item response theory (IRT) analysis, whereby the sample is first stratified into those with comorbid disorders and those without. Specifically, an IRT graded response (GRM) model was applied to the 4-item Likert scale (PHQ-4), and a depression latent ϴ metric was derived for each participant. A two-parameter IRT (2-PL) model was applied to the 12-item neuroticism binary response item scale (e.g., Bauermeister & Gallacher, 2019), and a post-estimation neuroticism latent ϴ metric was derived for each participant. There are several advantages of using ϴ over using a summated score in longitudinal studies. First, as a standardised score, the IRT ϴ provides a more stable and reliable construct over time. IRT is able to provide additional longitudinal information beyond just the score as it provides both within and between-item information across longitudinal waves (Rock, 2012). Also, ϴ is able to avoid overestimating the repeated measure variance and underestimating the between-person variance (Gorter, Fox, & Twisk, 2015). Moreover, it is suggested that the ϴ latent metric may be more sensitive to longitudinal change as it is computed along a continuum with multiple measurement points, compared to a single summated score which may contain as many as 12 measurement points. Second, it has also been suggested that using an IRT-based latent metric will produce smaller sample bias particularly in large cohorts (Lu, Thomas, & Zumbo, 2005). A positive ϴ suggests possession or endorsement of the trait and a negative ϴ

intuitively suggests absence or non-endorsement with most respondents recording values between – 3 and + 3 (e.g., De Mars, 2010). The sample is stratified accordingly, for instance, participants with positive neuroticism Θ are categorized into the anxiety group.

Step 2 – Identify the prediction pattern of mental disorders on cognition

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 (Ghazi et al., 2018; Pearlmutter, 1989).

Long short-term memory (LSTM) is an RNN architecture developed to deal with the exploding and vanishing gradient problem during backpropagation through time. By employing a constant error carousel: (1) a memory cell which maintains its state over arbitrary time intervals, and (2) three nonlinear gating units (input, forget and output) which regulate the information flow into and out of the cell, it is able to effectively capture long-term temporal dependencies (Gers & Schmidhuber, 2001; Gers, Schmidhuber, & Cummins, 1999; Ghazi et al., 2018; Greff, Srivastava, Koutník, Steunebrink, & Schmidhuber, 2017; Hochreiter & Schmidhuber, 1997). This study applies the commonly used LSTM setup described by Graves and Schmidhuber (Graves & Schmidhuber, 2005), which uses full gradient training and has been shown to perform reasonably well among other approaches in practice (Greff et al., 2017). 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 optimization 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 (Kingma & Ba, 2014). 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.

To evaluate whether depression and anxiety are important comorbid disorders of cognition, we first ran model (I) which only included covariates as features and was used as the reference model, then ran model (II) which included the neuroticism scales in addition to model (I), and model (III) which included the depression scales in addition to model (I). We also ran model (IV) which included diabetes and cardiovascular disease in addition to model (I), which was used to compare their prediction abilities with those of the mental health disorders. 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 demographic variables.

Results

Figure 1 summarises the percentage of participants with high and low IIV in each condition group and Table 2 reports the unadjusted relative risk under three time waves. For participants with anxiety, the percentage of having impaired cognition constantly increases through time from 35.09% to 42.15% for high IIV and 64.91% to 57.85% for low IIV; in contrast, increase with less magnitude is found for participants without anxiety from 36.10% to 40.03% for high IIV and 63.90% to 59.97% for low IIV. Likewise, for participants with depression, the percentage of having impaired cognition constantly increases through time from 35.49% to 41.67% for high IIV and from 64.51% to 58.33% for low IIV; in contrast, increase with less magnitude is found for participants without depression from 35.64% to 40.52% for high IIV and from 64.36% to 59.48% for low IIV.

The unadjusted relative risk increases from 0.972 at time wave 1 to 1.503 at time 3 for anxiety and 0.996 to 1.028 for depression[1]. These descriptive patterns suggest that comorbidity of anxiety and depression may lead to impaired cognition in the longer-term.

Table 3 reports the performance measures of the LSTM network using different features in predicting cognition, and Figure 2 represents the relational characteristics of the comorbidities on a ROC curve. It can be observed that the anxiety model achieved the best performance among all models, with an AUC of 0.68, followed by the depression model with an AUC of 0.64. The cardiovascular and diabetes model and the demographics model had relatively weak performance in predicting cognition, with an AUC of 0.60 and 0.57, respectively. Among all models, the anxiety model prospectively identified 62.89% of patients who eventually reached high IIV (i.e., sensitivity), and 53.27% of patients who kept 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 outperform the reference model (model that only includes covariates as features). In general, the neuroticism and depression scales helped improve the accuracy in capturing the pattern of reaction time, and therefore may be important predictors of impaired cognition. The results suggest increased risk of multiple mental health disorders occurring together, suggesting depression and anxiety are potential comorbid disorders of cognition.

Discussion

In this paper, a longitudinal analysis of the comorbid predictors of mental health disorders on cognitive decline was conducted using machine learning techniques. Using data

[1] In order to preserve a dataset as large as possible, when summarizing comorbid information, we include participants with no missing values on cognition and psychiatric disorders without considering covariates. The total number of available participants are 3501.

from UK Biobank, this study provides empirical evidence which suggests that anxiety and depression are important comorbid disorders of cognition but when cardiovascular disease, diabetes and other demographic predictors were included in the model these were not as significant on long-term effect. This suggests that poor mental health has a significant deleterious effect on long-term cognition, and may be considered a 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 general population large enough for identification of valid effects which might be implied across clinical and research settings.. Clinical-based samples have been criticized 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 the comorbid diagnoses of a patient (Cloninger, 2002). 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 (Angold, Costello, & Erkanli, 1999; Cloninger, 2002).

Second, this study investigates the psychiatric comorbid disorders of cognition in a comprehensive design in two stages. In the first stage, comorbidity information is summarized using a commonly used analytical approach. The variation of cognition for participants with or without psychiatric disorders through time shows an overview of the increased risk of mental health disorders and impaired cognition occurring together. In the second stage, a ML approach is 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 (Ghazi et al., 2018).

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.

The implications 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 both research and clinical settings.

Acknowledgements

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

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

Authors' Contributions

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

References

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *J Child Psychol Psychiatry, 40*(1), 57-87.

Arve, S., Tilvis, R. S., Lehtonen, A., Valvanne, J., & Sairanen, S. (1999). Coexistence of lowered mood and cognitive impairment of elderly people in five birth cohorts. *Aging (Milano), 11*(2), 90-95.

Bauermeister, S., & Bunce, D. (2015). Poorer mental health is associated with cognitive deficits in old age. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn, 22*(1), 95-105. doi:10.1080/13825585.2014.893554

Bauermeister, S., & Gallacher, J. E. (2019). A psychometric evaluation of the 12-item EPQ-R neuroticism scale in 384,183 UK Biobank participants using item response theory (IRT) *BioRxiv*. doi:https://doi.org/10.1101/741249

Bauermeister, S., Orton, C., Thompson, S., Barker, R. A., Bauermeister, J. R., Ben-Shlomo, Y., . . . Gallacher, J. E. (Preprint). Data Resource Profile: The Dementias Platform UK (DPUK) Data Portal. *BioRxiv*. doi:10.1101/582155

Biessels, G. J., Deary, I. J., & Ryan, C. M. (2008). Cognition and diabetes: a lifespan perspective. *Lancet Neurol, 7*(2), 184-190. doi:10.1016/S1474-4422(08)70021-8

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5-32.

Breiman, L. (2017). *Classification and regression trees*: Routledge.

Bunce, D., Batterham, P. J., Christensen, H., & Mackinnon, A. J. (2014). Causal associations between depression symptoms and cognition in a community-based cohort of older adults. *Am J Geriatr Psychiatry, 22*(12), 1583-1591. doi:10.1016/j.jagp.2014.01.004

Chang, Y.-J., Chen, L.-J., Chang, Y.-J., Chung, K.-P., & Lai, M.-S. (2012). Risk groups defined by Recursive Partitioning Analysis of patients with colorectal adenocarcinoma treated with colorectal resection. *BMC Medical Research Methodology, 12*(1), 2.

Chao, S. T., Koyfman, S. A., Woody, N., Angelov, L., Soeder, S. L., Reddy, C. A., . . . Suh, J. H. (2012). Recursive partitioning analysis index is predictive for overall survival in patients undergoing

spine stereotactic body radiation therapy for spinal metastases. *International Journal of Radiation Oncology* Biology* Physics, 82*(5), 1738-1743.

Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., . . . Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet Psychiatry, 3*(3), 243-250.

Cloninger, C. R. (2002). Implications of comorbidity for the classification of mental disorders: the need for a psychobiology of coherence. *Psychiatric diagnosis and classification, 79*, 106.

De Mars, C. (2010). *Item Response Theory*. New York, USA: Oxford University Press.

Dykiert, D., Der, G., Starr, J. M., & Deary, I. J. (2012). Age differences in intra-individual variability in simple and choice reaction time: systematic review and meta-analysis. *PloS one, 7*(10), e45759.

Eysenck, S. B. G., Eysenck, H. J., & Barrett, P. (1985). A Revised Version of the Psychoticism Scale. *Personality and Individual Differences, 6*(1), 21-29. doi:Doi 10.1016/0191-8869(85)90026-1

Feart, C., Samieri, C., & Barberger-Gateau, P. (2010). Mediterranean diet and cognitive function in older adults. *Curr Opin Clin Nutr Metab Care, 13*(1), 14-18. doi:10.1097/MCO.0b013e3283331fe4

Feinstein, A. R. (1970). The Pre-Therapeutic Classification of Co-Morbidity in Chronic Disease. *J Chronic Dis, 23*(7), 455-468.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1): Springer series in statistics New York.

Gers, F. A., & Schmidhuber, E. (2001). LSTM recurrent networks learn simple context-free and context-sensitive languages. *IEEE Transactions on Neural Networks, 12*(6), 1333-1340.

Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: Continual prediction with LSTM.

Ghazi, M. M., Nielsen, M., Pai, A., Cardoso, M. J., Modat, M., Ourselin, S., & Sørensen, L. (2018). Robust training of recurrent neural networks to handle missing data for disease progression modeling. *arXiv preprint arXiv:1808.05500*.

Gorter, R., Fox, J. P., & Twisk, J. W. (2015). Why item response theory should be used for longitudinal questionnaire data analysis in medical research. *BMC Med Res Methodol, 15*, 55. doi:10.1186/s12874-015-0050-x

Graves, A., & Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks, 18*(5-6), 602-610.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems, 28*(10), 2222-2232.

Hassing, L. B., Hofer, S. M., Nilsson, S. E., Berg, S., Pedersen, N. L., McClearn, G., & Johansson, B. (2004). Comorbid type 2 diabetes mellitus and hypertension exacerbates cognitive decline: evidence from a longitudinal study. *Age Ageing, 33*(4), 355-361. doi:10.1093/ageing/afh100

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735-1780.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kraemer, H. C. (1995). Statistical issues in assessing comorbidity. *Stat Med, 14*(8), 721-733.

Kroenke, K., Spitzer, R. L., Williams, J. B., & Lowe, B. (2009). An ultra-brief screening scale for anxiety and depression: the PHQ-4. *Psychosomatics, 50*(6), 613-621. doi:10.1176/appi.psy.50.6.613

Lu, I. R. R., Thomas, D. R., & Zumbo, B. D. (2005). Embedding IRT in Structural Equation Models: A Comparison With Regression Based on IRT Scores. *Structural Equation Modeling, 12*(2), 263-277.

Lyall, D. M., Cullen, B., Allerhand, M., Smith, D. J., Mackay, D., Evans, J., . . . Deary, I. J. (2016). Cognitive test scores in UK Biobank: data reduction in 480,416 participants and longitudinal stability in 20,346 participants. *PloS one, 11*(4), e0154222.

Mathew, A. R., Hogarth, L., Leventhal, A. M., Cook, J. W., & Hitsman, B. (2017). Cigarette smoking and depression comorbidity: systematic review and proposed theoretical model. *Addiction, 112*(3), 401-412. doi:10.1111/add.13604

Meghani, S. H., Buck, H. G., Dickson, V. V., Hammer, M. J., Rabelo-Silva, E. R., Clark, R., & Naylor, M. D. (2013). The conceptualization and measurement of comorbidity: a review of the interprofessional discourse. *Nurs Res Pract, 2013*, 192782. doi:10.1155/2013/192782

Munshi, U. M., & Chaudhuri, B. B. (2011). *Multimedia information extraction and digital heritage preservation* (Vol. 10): World Scientific.

Panza, F., Frisardi, V., Capurso, C., D'Introno, A., Colacicco, A. M., Imbimbo, B. P., . . . Solfrizzi, V. (2010). Late-life depression, mild cognitive impairment, and dementia: possible continuum? *Am J Geriatr Psychiatry, 18*(2), 98-116. doi:10.1097/JGP.0b013e3181b0fa13

Pearlmutter, B. A. (1989). Learning state space trajectories in recurrent neural networks. *Neural Computation, 1*(2), 263-269.

Piumatti, G., Moore, S., Berridge, D., Sarkar, C., & Gallacher, J. (2018). The relationship between alcohol use and long-term cognitive decline in middle and late life: a longitudinal analysis using UK Biobank. *J Public Health (Oxf), 40*(2), 313-314. doi:10.1093/pubmed/fdy032

Ripley, B. D. (2007). *Pattern recognition and neural networks*: Cambridge university press.

Rock, D. (2012). Modeling change in large-scale longitudinal studies of eduational growth: four decades of contributions to the assessment of education growth. *ETS R&D Scientific and Policy Contributions Series, ETS SPC-12-01*.

Roe, M., Pinchen, H., Church, S., & Finglas, P. (2015). McCance and Widdowson's the composition of foods seventh summary edition and updated composition of foods integrated dataset. *Nutrition Bulletin, 40*(1), 36-39.

Shouval, R., Bondi, O., Mishan, H., Shimoni, A., Unger, R., & Nagler, A. (2014). Application of machine learning algorithms for clinical predictive modeling: a data-mining approach in SCT. *Bone marrow transplantation, 49*(3), 332.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Landray, M. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine, 12*(3), e1001779.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., . . . Collins, R. (2015). UK Biobank: An Open Access Resource for Indentifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *12*(3). doi:10.1371/journal.pmed.1001779

van Loo, H. M., Cai, T., Gruber, M. J., Li, J., de Jonge, P., Petukhova, M., . . . Wardenaar, K. J. (2014). Major depressive disorder subtypes to predict long ‐ term course. *Depression and anxiety, 31*(9), 765-777.

Wardenaar, K. J., van Loo, H. M., Cai, T., Fava, M., Gruber, M. J., Li, J., . . . Rose, S. (2014). The effects of co-morbidity in defining major depression subtypes associated with long-term course and severity. *Psychological medicine, 44*(15), 3289-3302.

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									
"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 half	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 half	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 half	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 half	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
Other	0.44	75.91	54.56	0.44	72.50	47.16	0.42	79.61	51.72

BMI									
Normal weight	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,£30,999]	0.22	76.05	44.03	0.27	74.34	47.62	0.27	78.97	47.54
(£31,000,£51,999]	0.30	72.51	47.30	0.30	73.71	50.86	0.30	69.63	45.74
(£52,000,£100,000]	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									
Yes	0.95	72.73	50.16	0.96	72.57	48.36	0.97	75.62	51.64
No	0.05	63.09	34.48	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

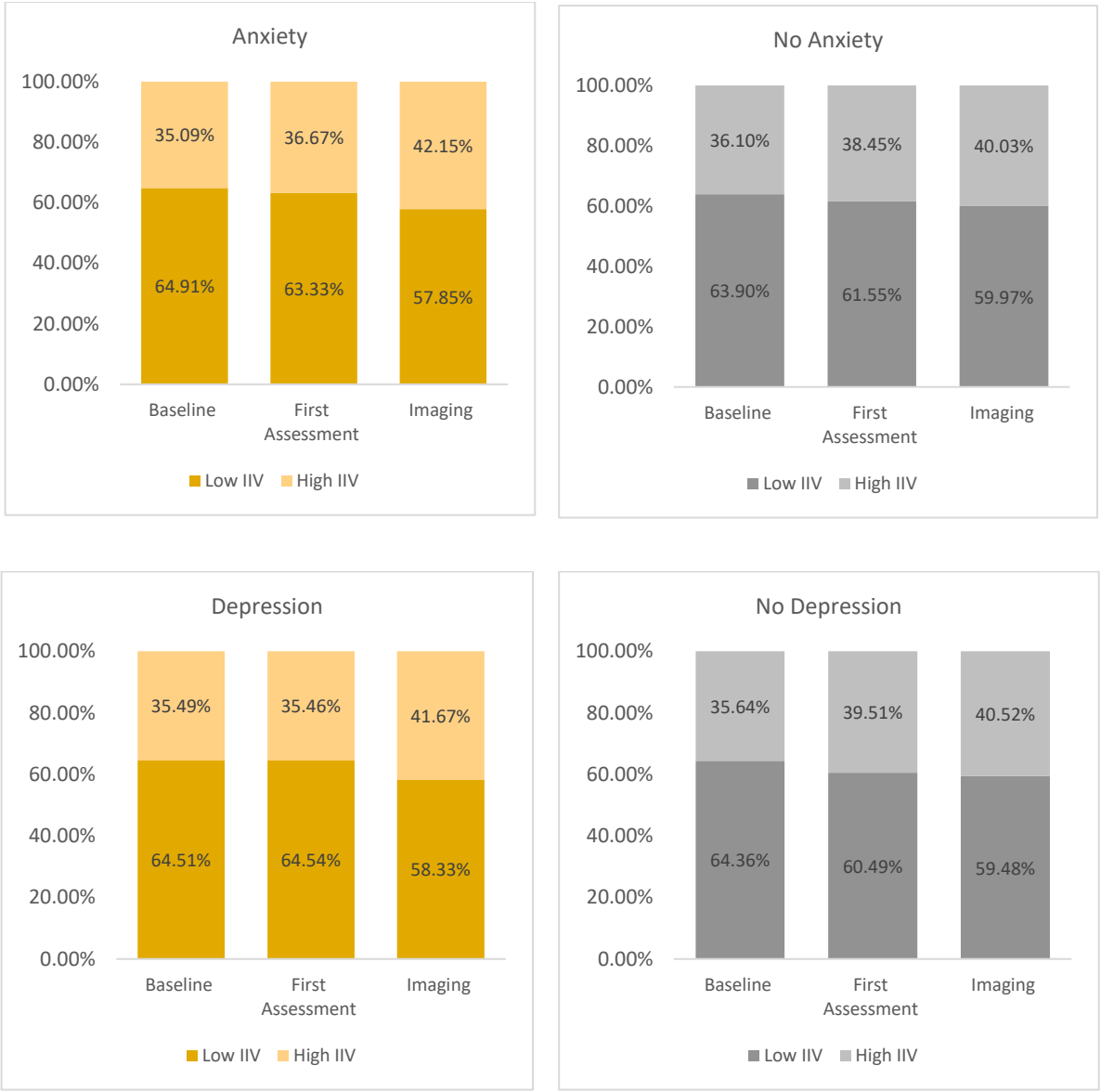

Figure 1 Cognition distribution

Table 2 Unadjusted relative risk

	Time wave 1	Time wave 2	Time wave 3
Anxiety	0.972	0.954	1.053
Depression	0.996	0.898	1.028

Table 3 Model performance

	Anxiety	Depression	Cardiovascular & Diabetes	Demographics
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.

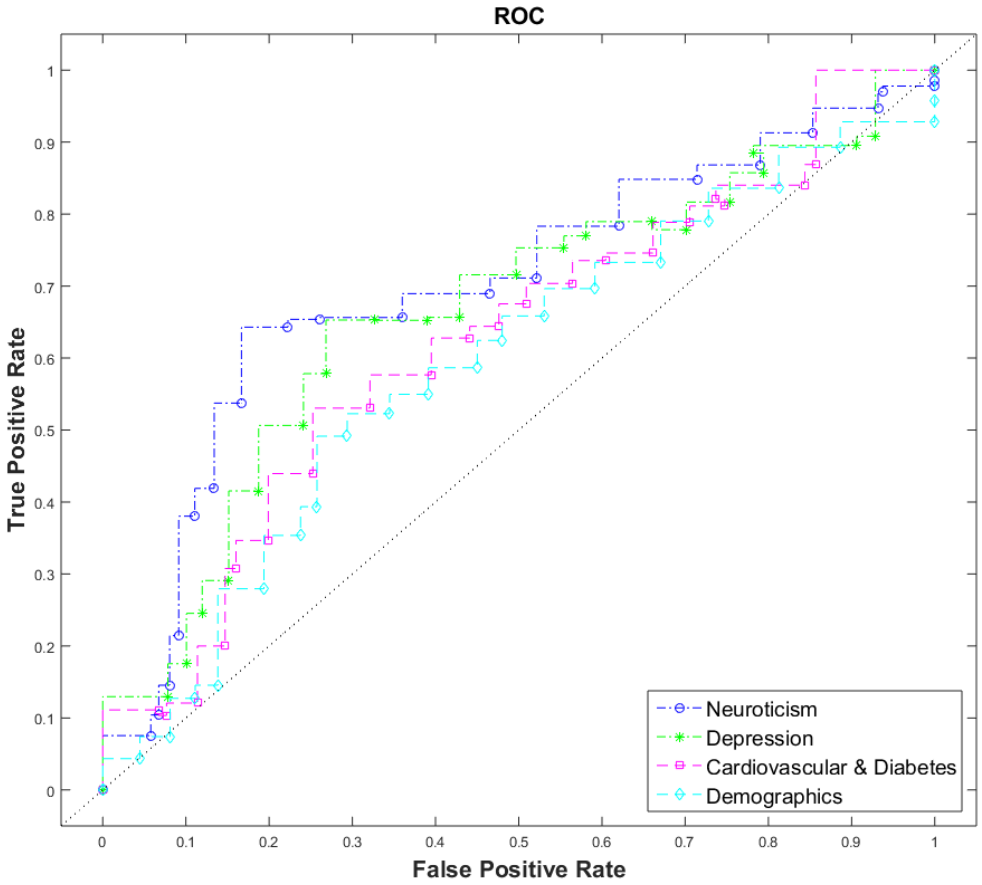

Figure 2 ROC