# Development of Digitally Obtainable 10-Year Risk Scores for Depression and Anxiety in the General Population

Davide Morelli[1,2], Nikola Dolezalova[1], Sonia Ponzo[1], Michele Colombo[1] and David Plans[1,3,4]*

[1] Huma Therapeutics Ltd., London, United Kingdom, [2] Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, United Kingdom, [3] Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom, [4] Initiative in the Digital Economy at Exeter (INDEX) Group, Department of Science, Innovation, Technology, and Entrepreneurship, University of Exeter, Exeter, United Kingdom

The burden of depression and anxiety in the world is rising. Identification of individuals at increased risk of developing these conditions would help to target them for prevention and ultimately reduce the healthcare burden. We developed a 10-year predictive algorithm for depression and anxiety using the full cohort of over 400,000 UK Biobank (UKB) participants without pre-existing depression or anxiety using digitally obtainable information. From the initial 167 variables selected from UKB, processed into 429 features, iterative backward elimination using Cox proportional hazards model was performed to select predictors which account for the majority of its predictive capability. Baseline and reduced models were then trained for depression and anxiety using both Cox and DeepSurv, a deep neural network approach to survival analysis. The baseline Cox model achieved concordance of 0.7772 and 0.7720 on the validation dataset for depression and anxiety, respectively. For the DeepSurv model, respective concordance indices were 0.7810 and 0.7728. After feature selection, the depression model contained 39 predictors and the concordance index was 0.7769 for Cox and 0.7772 for DeepSurv. The reduced anxiety model, with 53 predictors, achieved concordance of 0.7699 for Cox and 0.7710 for DeepSurv. The final models showed good discrimination and calibration in the test datasets. We developed predictive risk scores with high discrimination for depression and anxiety using the UKB cohort, incorporating predictors which are easily obtainable *via* smartphone. If deployed in a digital solution, it would allow individuals to track their risk, as well as provide some pointers to how to decrease it through lifestyle changes.

Keywords: depression, machine learning, anxiety, prediction model, risk scores

## INTRODUCTION

Global prevalence of depression was estimated to be 280 million (1) in 2019. By 2030, depression is expected to be the second-largest contributor to worldwide loss of years of healthy life because of death or disability (2). Highly comorbid with depression, anxiety disorders globally are estimated to affect 301 million individuals (1). NICE guidelines currently recommend the use of validated questionnaires [e.g., PHQ-9, Patient Health Questionnaire (3); HADS, Hospital Anxiety and

Depression Scale (4) and BDI, Beck Depression Inventory (5) for depression; GAD-2 or GAD-7 (6) for anxiety disorders] to diagnose patients and classify the severity of their symptoms (7). Whilst tools based on patients' self-reported feelings and mood changes are invaluable to track the progression of the disorder, multifactorial models including well-established risk factors are needed to successfully manage the disorder in the long-term. Predictive scores can be used to effectively identify patients at highest risk of developing depression or anxiety and enroll them in preventative pathways, thus, minimizing relapses and lowering the burden of the disease (8).

A comprehensive review of existing predictive scores is outside of the scope of this paper and can be found elsewhere (9). There are several scores for specific populations at risk of depression, e.g., adolescents (10), elderly (11), traumatic head injury (12) or stroke patients (13), patients with diabetes (14), or immune-mediated inflammatory disorders (15). For the general population, the most widely used depression risk score is the PredictD score, developed using patient data from six European countries and externally validated on a population from Chile (16). The original score contains 10 risk factors (age, sex, country, education level, personal and family history of depression, physical and mental health disturbances, difficulties at work, and experience of discrimination). Other country-specific scores have been developed after PredictD to better account for cultural and socio-economic differences (17–20), but little research has been conducted to develop risk scores aimed at predicting onset of generalized anxiety and panic disorders in the general population. The PredictA score was developed using the same dataset described above for identifying factors predicting depression, and it includes sex, age, lifetime depression, family history of psychological difficulties, physical health, and mental health disturbances, unsupported difficulties in paid and/or unpaid work, country of residence and time of follow-up (21). Given the high comorbidity between anxiety and depressive disorders, it has also been suggested that, on top of disorder-specific risk factors, a set of common underlying risk factors for both disorders may exist (22).

The majority of the published risk scores for depression and anxiety relate to short term predictions (between 6 and 24 months) that mainly involve non-modifiable factors (e.g., family history). Given the impact of recurring episodes and lifetime duration on the progression of the disorders (23, 24), early identification of individuals at risk of depression or anxiety would be beneficial in devising effective preventative and therapeutic pathways. Evidence suggests that the risk of depression and anxiety can be decreased by modifying certain lifestyle factors, such as having a balanced diet and performing physical activity (25, 26) or smoking cessation (27). With an increased availability and need for telematics solutions in healthcare due to the COVID-19 pandemic (28–30), risk scores evidencing modifiable lifestyle changes have the potential to be of wide benefit.

The UK Biobank (UKB) is a prospective study of over half a million UK participants, recruited between 2006 and 2010. Available data includes primary care and hospital inpatient records, results of touchscreen questionnaires and verbal interviews about lifestyle, pre-existing conditions and family history, as well as a comprehensive battery of medical tests,

imaging and physical assessments. While predictive models have been developed for many common diseases, such as cardiovascular diseases or diabetes (31, 32), no long-term predictive models have been derived for depression and anxiety using the UKB cohort. This dataset contains information about many known predictors of severe mental illness, including modifiable lifestyle factors and other digitally-obtainable data such as comorbidities, socioeconomic factors, or early-life events (33, 34).

The aim of this study is to devise a model of potential risk factors for depression and anxiety in the long term (10+ years) using the UKB data and with a focus on behavioral and health indices that can be potentially tracked by means of a remote digital solution.

# METHODS

## Data Source and Study Design

The use of data for this study was approved by UKB, under the project title "Validation and comparative analysis of novel prediction models focused on modifiable lifestyle factors for the risks of common, preventable diseases and all-cause mortality: a cohort study" (application number 55668).

The aim of this study was to model the risk of being diagnosed with depression or anxiety for the first time over the next 10 years. The outcomes were defined as an occurrence of a depressive episode (ICD10 code F32) or an anxiety disorder (ICD10 code F41), respectively, after the date of assessment. These were derived from the UKB First Occurrences fields 130894 and 130906 which combine data from primary care and hospital inpatient diagnoses with conditions self-reported at the time of assessment. Primary and secondary care data contains the exact date of each diagnosis. Results from diagnostic mental health questionnaires were not used in the definition of the outcome because the participants were not systematically screened over the follow-up period.

Potential predictor variables include results from the touchscreen questionnaire administered at the initial assessment (instance 0) and pre-existing illnesses diagnosed by the time of assessment. Only fields which could be collected *via* a smartphone app (i.e., using the sensors available on smartphones or *via* user's input), collected at or before the time of the initial assessment were included. The final set of potential predictors includes demographic data (age, sex, ethnicity), socioeconomic status (income, qualifications), physio-metric data (body mass index, height, weight, etc.,), family history (illnesses of parents or siblings), medical history, lifestyle characteristics (physical activity, diet, sleep habits), mental health history, moods and overall perceived well-being (satisfaction with life, mood swings, feelings of worry, loneliness, etc.,). Continuous variables with a high proportion of missing values (>10%) were excluded, as well as fields which could cause label leakage (**Supplementary Table 1**).

## Data Preparation

Participants were excluded from the study if they were diagnosed with the corresponding outcome condition prior to the date of assessment. Survival time was measured as years from the date

of assessment (instance 0 in the UK BioBank dataset) until the date of depression/anxiety diagnosis, or in participants who were not diagnosed, data was right-censored at the data extraction date (30th September 2020), date of death or date when they were lost to follow-up.

The binary variable "any_mental_issue" was derived from the set of First Occurrences fields for ICD10 diagnoses in the "Mental and behavioral disorders" category (F00–F72, excluding F32 and F33 in the Depression model and F40 and F41 in the Anxiety model). Only dates prior to the date of assessment were considered for this variable.

For CoxPH, all categorical features were one-hot encoded, followed by exclusion of categories containing <0.1% items; for DeepSurv, the categorical features were encoded using one-hot, target-encoding, and weight of evidence (the encoding being one of the parameters), followed by exclusion of categories containing <0.1% items.

For six diet-related features, values "Do not know" and "Prefer not to answer" were substituted with mean. Continuous features were centered and scaled to unit variance (details in **Supplementary Table 2**). Finally, participants with any remaining missing continuous features were excluded (characteristics of excluded participants in **Supplementary Tables 3, 4**).

Test set (25%) was set aside for internal validation using a stratified train-test split (preserving the ratios seen in the binary outcome field). From the train set, a further 25% was set aside as a validation set and used for feature selection and the optimisation of DeepSurv parameters. For the final models after feature selection, the train and validation dataset were combined for training, followed by evaluation on the test dataset (**Supplementary Figure 1**).

## Baseline Model and Feature Selection

First, a Cox-Proportional Hazards (CPH) model, implemented in the Python lifelines library (35), was trained using the full set of features. The number of features was then reduced to decrease the risk of overfitting, improve the explainability of the model and to narrow down the number of inputs from users in a potential digital solution. As a first step, univariate Cox analysis was performed for each feature and those with $p$-value over 0.1 were excluded. Then, an iterative backward elimination algorithm was used to get the final set of features. In short, in every round of elimination, the CPH model was retrained without a set of features with the highest $p$-value. If the concordance index evaluated on the validation dataset decreased by more than 0.001, features were kept for an additional round testing elimination of a smaller number of features. Elimination of each remaining single feature was tested before the decision to keep it in the reduced model.

The reduced model was then trained using this final set of features on joined train and validation dataset and its performance evaluated on the unseen test cohort. Finally, this model was reviewed and variables with problematic clinical explanation were manually removed from the model.

## Deep Survival Analysis

The next model we tested was the Cox proportional hazards deep neural network (DeepSurv), using an implementation in the "pycox" package (36) based on the deep learning library PyTorch (37). Details of all libraries and their versions used in this study can be found in **Supplementary Table 5**.

Using either the full or reduced set of features, we searched the hyperparameter space using a set of parameters described in **Supplementary Table 6**. This was done using a Tree-Structured Parzen Estimator algorithm (38) from the Optuna Library (39). In total, 500 configurations were tested for both depression and anxiety, separately for baseline and reduced models, each evaluated on the validation dataset. Feedforward neural networks deep up to three hidden layers have been tested. Classic Stochastic Gradient Descent algorithms with Momentum (40) and Adam (41) with optimal learning rate estimation were used for training. The best combination of hyperparameters was selected separately for the baseline and reduced models for depression and anxiety. The performance was then compared to a neural network with the same hyperparameters but using only the reduced set of features selected by the backwards elimination using Cox classifier.
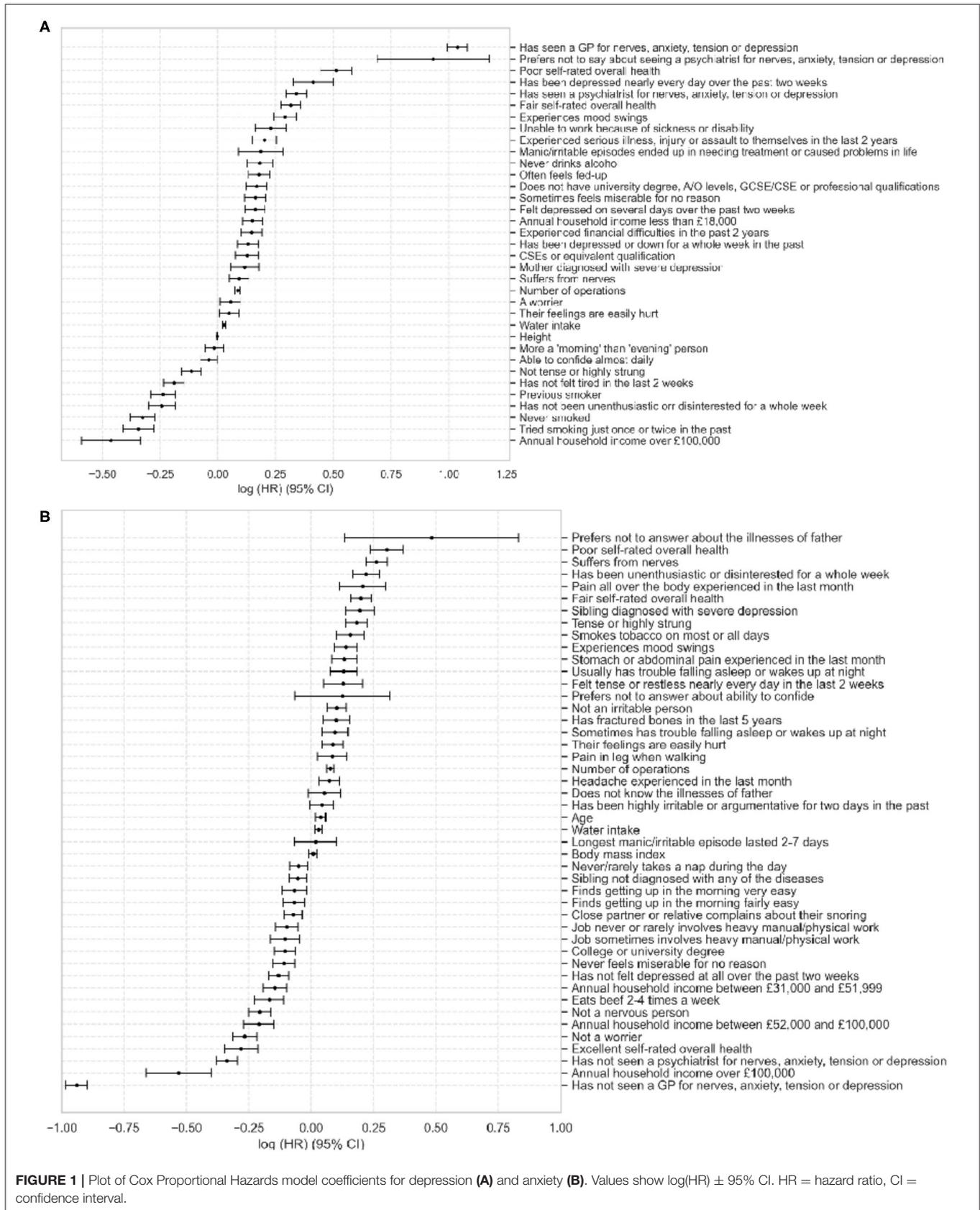
## Statistical Analysis

Results from the analysis of demographic characteristics show participant numbers and percentages of total for categorical/ordinal variables, or medians and quartiles (Q1 and Q3) for continuous variables. Statistical comparisons were performed using the Chi-squared test for categorical/ordinal and Kruskal-Wallis test for continuous variables.

C-index was used as the metric for all models, with 95% confidence intervals calculated using the percentile bootstrap resampling method (50 resampling rounds). Where detailed analysis of the results of CPH models is provided, logarithm of hazard ratios/log(HR) with 95% confidence intervals (CIs) are shown. $P$-values test the null hypothesis that the coefficient of each variable is equal to zero and significance level was set

**TABLE 1** | Results of the Cox Proportional Hazards model for depression and anxiety.

| | # Features | C-index (train) | C-index (validation) |
|---|---|---|---|
| **Before feature selection** | | | |
| Depression | 429 | 0.7891 [0.7883, 0.7895] | 0.7901 [0.7883, 0.7919] |
| Anxiety | 429 | 0.7739 [0.7731, 0.7747] | 0.7650 [0.7634, 0.7665] |
| **After feature selection** | | | |
| Depression | 35 | 0.7832 [0.7830, 0.7835] | 0.7796 [0.7793, 0.7802] |
| Anxiety | 46 | 0.7682 [0.7680, 0.7684] | 0.7668 [0.7662, 0.7674] |

*Mean concordance index (C-index) is shown for training (train+validation) and test dataset. 95% confidence intervals are shown in square brackets.*

FIGURE 1 | Plot of Cox Proportional Hazards model coefficients for depression **(A)** and anxiety **(B)**. Values show log(HR) ± 95% CI. HR = hazard ratio, CI = confidence interval.

to 0.05. Calibration was evaluated at the 10-year time point using calibration plots and the Integrated Calibration Index (ICI), which is a mean weighted difference between observed and predicted probabilities, implemented in the Python lifelines library (35).

# RESULTS

## Study Population

From the initial set of 502,488 participants in the UK Biobank, 40,367 had pre-existing depression and 11,296 pre-existing anxiety. These participants were excluded from the respective datasets. Further participants were excluded due to missing values in some continuous or ordinal variables. In the depression dataset, of the remaining 448,733 participants, 16,507 (3.68%) developed depression after assessment. For the anxiety dataset, it was 17,830 (3.74%) out of 477,100 participants diagnosed with anxiety after assessment. Details on distribution of the outcomes in the train, validation and test datasets can be found in **Supplementary Figure 2**. There were no significant differences in any features between the train + validation and test datasets (**Supplementary Tables 7, 8**). Median follow-up time of 11.2 years and maximum follow-up time of 13.8 years were the same for depression and anxiety. Distribution of the durations to depression or anxiety development can be found in **Supplementary Figures 2A,B**.

The dataset used in this study contains 46% men and 54% women, aged 56 ± 8 (range 38–73) at the time of the initial assessment (**Supplementary Figures 3C,D**). The ethnic background of the participants was 94.5% white, 2.1% Asian, 1.5% black, and 1.9% other or unknown. Summary of the variables in the final model for depression and anxiety is presented in **Supplementary Tables 9, 10**, respectively.

## Cox Model and Feature Selection

The initially selected 167 UK Biobank variables were pre-processed into 429 feature columns for depression/anxiety (**Supplementary Table 2**). These were used to build a baseline Cox proportional hazards model for each of the two outcomes, achieving a concordance index of 0.7901 for depression and 0.7650 for anxiety in the validation cohort (**Table 1**).

After feature selection, the number of predictors for depression was narrowed down to 39 and for anxiety to 53. After manual review (excluded features and their coefficients can be found in **Supplementary Table 11**), the final number of predictors was 35 and 46, respectively. Decrease in number of predictors was accompanied by a slight decrease of the concordance index of depression model to 0.7796, the anxiety model maintained its performance at 0.7668 (**Table 1**). The predictors in the final models, along with their coefficients and confidence intervals are shown in **Figure 1** and **Supplementary Tables 12, 13**.

The top three risk factors in the depression model were seeing a GP for nerves, anxiety, tension, or depression, not wanting to talk about seeing a psychiatrist about these conditions and poor overall health. The most protective factors were annual household income over £100,000 and being a non-smoker. For

**TABLE 2 |** Summary of the best-performing DeepSurv models for depression and anxiety.

|  | Before feature selection | | After feature selection | |
| --- | --- | --- | --- | --- |
|  | # Features | C-index | # Features | C-index |
| Depression | 523 | 0.7878 [0.7866, 0.7894] | 35 | 0.7863 [0.7723, 0.7873] |
| Anxiety | 529 | 0.7728 [0.7714, 0.7737] | 46 | 0.7710 [0.7710, 0.7719] |

*Mean C-index calculated on the test dataset, before and after feature selection. 95% confidence intervals are shown in square brackets.*

anxiety, the top risk factors were not wanting to talk about father's illnesses, poor self-rated health and suffering from health, the protective factors were never visiting a GP or psychiatrist for nerves, anxiety, tension, or depression and annual household income over £100,000.

The mean predicted 10-year risk of developing depression was 3.30% (95% CI 3.27–3.32), the mean observed risk was 3.21% (95% CI 3.18–3.23). The anxiety model predicted an average risk of 3.26% (95% CI 3.23–3.28), while the observed probability was 3.18% (95% CI 3.16–3.20). The Cox models showed good calibration, particularly for the low probabilities which were abundant in the population, with slightly larger errors for the higher probabilities which were sparsely represented. The Integrated Calibration Index (ICI) of 0.0009 for depression and 0.0013 for anxiety (**Supplementary Figure 4**).

## Machine Learning Models

The hyperparameter space for the DeepSurv model was explored by running 500 combinations, best of which achieved a concordance index of 0.7878 (95% CI 0.7866–0.7894) for depression and 0.7728 (95% CI 0.7714–0.7737) for anxiety using the full set of variables (**Table 2**). Using the reduced set of features after feature selection, the depression model showed a concordance of 0.7863 (95% CI 0.7723–0.7873) and the anxiety model concordance of 0.7710 (95% CI 0.7710–0.7719) (**Table 2**). The details of the hyperparameters for the best-performing models can be found in **Supplementary Table 6**.

## Comparison of the Depression and Anxiety Models

Of the 35 features in the depression model and 46 features in the anxiety model, 8 features appeared in both models (**Figure 2**). These include Poor or fair self-rated health, experience of mood swings, suffering from nerves, number of operations, having their feelings easily hurt, and water intake. An annual household income over £100,000 had negative coefficients in both models.

**FIGURE 2 |** Visual comparison of the reduced models for depression and anxiety. Features are positioned on the y axes in the order of descending coefficients but the distance between points is not proportional to the difference between coefficients. Connecting lines are shown in red (for common factors with a higher coefficient in the depression model) or green (higher coefficient in anxiety model).

## DISCUSSION

### Key Findings

The aim of the current study was to build prediction models for depression and anxiety, with a specific focus on factors linked to digitally-obtainable data. Using a data-driven approach to feature selection and model optimisation, we trained models for prediction of depression and anxiety using both traditional

statistical and machine learning methods. The best-performing model for depression achieved a concordance index of 0.7863, with similar performance to other similar depression risk scores. For comparison, the current golden standard score PredictD achieved a concordance index of 0.790 using stepwise logistic regression [16]. Rosellini et al. [19] used an ensemble machine learning algorithm with a concordance of 0.757, whereas, Wang et al. [20] developed sex-specific logistic regression models with

concordance index of 0.795 for men and 0.767 for women. Our DeepSurv model for anxiety shows a concordance of 0.771, compared to 0.752 in the PredictA study (21). It is important to note that participants in our study were followed for more than 10 years while the prediction horizon in the other studies was shorter (1–4 years). Among the most similar studies developed on the UKB dataset, Zhou et al. achieved a concordance index of 0.778 for prediction of depressive moods using neuroimaging and questionnaire data and Sarris et al. (42) in their study of lifestyle factors associated with frequency of depressive moods built an ordinal logistic regression model but did not report its performance. To the best of our knowledge, this is the first study which used the UKB cohort to develop long-term prognostic risk scores for both depression and anxiety.

The DeepSurv model performed comparably to the Cox model. While it has the capability to capture complex non-linear relationships between factors, this fact could point to a linear relationship of most variables included in the model. An alternative explanation could be based on the requirement of deep learning algorithms for very large datasets to model very complex problems. It is possible that the problem in hand is too complex to be modeled with the provided number of training examples and deep learning thus cannot overperform traditional models. For the use of this score in a digital healthcare setting, interpretability of the score is key. With DeepSurv, similar to many other black-box machine learning models, the direction and scale of each feature's contribution to the overall risk may not be easily understandable. From this perspective, Cox model coefficients provide more intuitive understanding of how each feature could be changed to decrease the predicted risk, an information which could potentially motivate the user toward the right lifestyle changes.

Our prediction model includes many traditional risk factors for mental health illness, including smoking, alcohol consumption, employment status, overall health status, sleeping disturbances, social functioning, or education and income level (33). Interestingly, while age, sex, and ethnicity are among the risk factors in the existing predictive scores (16, 18, 21), they were eliminated from our models during the feature selection, with the exception of age in the anxiety model. We assume that the limited representation of all age and ethnic groups in the UKB dataset may be the reason some of these were not identified as important predictors. Nevertheless, the developed model provides a good discrimination capability for populations comprising predominantly white individuals and within the age range of ~40–70. Its performance on other demographics must be further investigated.

From the comparison of the final set of predictors in the depression and anxiety models it is clear that some pre-existing mental health conditions adversely affect the risk: seeing or not seeing a GP/psychiatrist for nerves, anxiety, tension or depression in the past is among the most important features in both models. Other factors featuring in both depression and anxiety models include the self-reported overall health rating, and associated variable of

number of operations. Our analysis also showed that household income or certain personality traits (being a moody person or easily hurt) are common risk factors for depression and anxiety.

Some of the predictors in our final model are dynamic factors, changing over time (e.g., smoking status, alcohol consumption, feelings or tiredness over the past 2-week period). Regular reassessment of the score would therefore be beneficial and could be aided by deployment of the score as a digital solution. Such application could also provide personalized recommendations on how to change one's lifestyle to decrease the risk of development of depression or anxiety.

## Study Limitations

Among the limitations of this study is the selection bias in the UKB cohort used to train the prediction models. It has been reported that the participants of UKB are on average healthier and come from less deprived areas (43). Notably, the representation of ethnicities in the dataset deviates from the general population, with over 94% participants being white, compared to 86% in the general population according to the latest UK census. Therefore, caution should be exercised when making predictions for individuals with a minority background, as they were underrepresented in the training dataset. This is an important limitation because they may be disproportionately affected by severe mental illnesses (44). The study also included only participants with a limited age range: 37–73 years at the time of assessment. Therefore, until these models have been validated on an appropriate cohort extending this range, predictions for users younger or older than this should be interpreted with reserve. The results of the internal validation performed in this study using an unseen test cohort does not point to overfitting but external validation will be necessary to confirm this.

Missing records or methods of variable encoding might also be introducing potential bias into this study. The goal of the study was to develop two separate models for anxiety and depression, we therefore, decided to keep diagnosis of anxiety as a predictor for the depression model, and vice versa, because of their high comorbidity. This setup would not be suitable if the goal were to build a single model predicting either depression or anxiety or both conditions simultaneously.

## Implications of Our Findings

This tool is intended to be used as a digital solution for dynamic tracking of individuals' risk of developing depression or anxiety. Evidence suggests that most individuals value knowing their risk of depression, especially if ways of prevention are also indicated (45, 46). Providing a personalized depression/anxiety risk score therefore appears as a safe preventative strategy with demonstrated benefits (46).

External validation of the developed model on data collected from another population will be necessary to be able to extend these findings to populations from geographies other than the United Kingdom. If these populations show significant

differences in demographic characteristics, such as ethnicity, causing the accuracy of the predictions to drop, the model can be re-calibrated if a suitable dataset exists for this population.

Additional improvements to the prediction models for mental health would be an expansion of the input space with other dynamic parameters, such as activity measures. Raw accelerometer data is available for a proportion of UK Biobank participants, and it could be used to provide an objective view on individuals' level of activity, sedentary time and sleep, with known associations with the risk of development of mental illness (33, 42, 47). Because these would be obtainable *via* wearable devices, they bring the potential of truly dynamic monitoring of an individual's risk.

## CONCLUSION

In summary, we present algorithms for prediction of depression and anxiety, developed using a large UK cohort of individuals followed for over 10 years. All factors in our models are easily acquirable *via* smartphone devices and thus can be used to support development of preventative digital solutions for mental health.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: the dataset analyzed in this study has been provided by the UK BioBank: https://www.ukbiobank.ac.uk/.

## AUTHOR CONTRIBUTIONS

DM designed the study. MC assisted with implementation of the used methods. DM and ND trained the models and performed data analysis. DM, ND, SP, and DP wrote the manuscript. All authors revised the manuscript.

## FUNDING

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2021.689026/full#supplementary-material

## REFERENCES

1. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet.* (2020) 396:1204–22. doi: 10.1016/S0140-6736(20)30925-9

2. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* (2006) 3:e442. doi: 10.1371/journal.pmed.0030442

3. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* (2001) 16:606–13. doi: 10.1046/j.1525-1497.2001.016009606.x

4. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* (1983) 67:361–70. doi: 10.1111/j.1600-0447.1983.tb09716.x

5. Beck AT, Steer RA, Ball R, Ranieri WF. Comparison of beck depression inventories-IA and -II in psychiatric outpatients. *J Pers Assess.* (1996) 67:588–97. doi: 10.1207/s15327752jpa6703_13

6. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* (2006) 166:1092–7. doi: 10.1001/archinte.166.10.1092

7. National Institute for Clinical Excellence and others. *Depression in adults. Quality standard [QS8].* London: NICE (2011).

8. Fernández A, Mendive JM, Conejo-Cerón S, Moreno-Peral P, King M, Nazareth I, et al. A personalized intervention to prevent depression in primary care: cost-effectiveness study nested into a clustered randomized trial. *BMC Med.* (2018) 16:28. doi: 10.1186/s12916-018-1005-y

9. Bernardini F, Attademo L, Cleary SD, Luther C, Shim RS, Quartesan R, et al. Risk prediction models in psychiatry: toward a new frontier for the prevention of mental illnesses. *J Clin Psychiatry.* (2017) 78:572–83. doi: 10.4088/JCP.15r10003

10. Van Voorhees BW, Paunesku D, Gollan J, Kuwabara S, Reinecke M, Basu A. Predicting future risk of depressive episode in adolescents: the Chicago Adolescent Depression Risk Assessment (CADRA). *Ann Fam Med.* (2008) 6:503–11. doi: 10.1370/afm.887

11. Shin K-R, Jung D, Jo I, Kang Y. Depression among community-dwelling older adults in Korea: a prediction model of depression. *Arch Psychiatr Nurs.* (2009) 23:50–7. doi: 10.1016/j.apnu.2008.03.001

12. Levin HS, McCauley SR, Josic CP, Boake C, Brown SA, Goodman HS, et al. Predicting depression following mild traumatic brain injury. *Arch Gen Psychiatry.* (2005) 62:523–8. doi: 10.1001/archpsyc.62.5.523

13. Christensen MC, Mayer SA, Ferran J-M, Kissela B. Depressed mood after intracerebral hemorrhage: the FAST trial. *Cerebrovasc Dis.* (2009) 27:353–60. doi: 10.1159/000202012

14. Jin H, Wu S. Use of patient-reported data to match depression screening intervals with depression risk profiles in primary care patients with diabetes: development and validation of prediction models for major depression. *JMIR Format Res.* (2019) 3:e13610. doi: 10.2196/13610

15. Tennenhouse LG, Marrie RA, Bernstein CN, Lix LM. Machine-learning models for depression and anxiety in individuals with immune-mediated inflammatory disease. *J Psychosom Res.* (2020) 134:110126. doi: 10.1016/j.jpsychores.2020.110126

16. King M, Walker C, Levy G, Bottomley C, Royston P, Weich S, et al. Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study. *Arch Gen Psychiatry.* (2008) 65:1368–76. doi: 10.1001/archpsyc.65.12.1368

17. Bellón JÁ, de Dios Luna J, King M, Moreno-Küstner B, Nazareth I, Montón-Franco C, et al. Predicting the onset of major depression in primary care: international validation of a risk prediction algorithm from Spain. *Psychol Med.* (2011) 41:2075. doi: 10.1017/S0033291711000468

18. Wang J, Sareen J, Patten S, Bolton J, Schmitz N, Birney A. A prediction algorithm for first onset of major depression in the general population: development and validation. *J Epidemiol Community Health.* (2014) 68:418–24. doi: 10.1136/jech-2013-202845

19. Rosellini AJ, Liu S, Anderson GN, Sbi S, Tung ES, Knyazhanskaya E. Developing algorithms to predict adult onset internalizing disorders: an ensemble learning approach. *J Psychiatr Res.* (2020) 121:189–96. doi: 10.1016/j.jpsychires.2019.12.006

20. Wang JL, Manuel D, Williams J, Schmitz N, Gilmour H, Patten S, et al. Development and validation of prediction algorithms for major depressive episode in the general population. *J Affect Disord.* (2013) 151:39–45. doi: 10.1016/j.jad.2013.05.045

21. King M, Bottomley C, Bellón-Saameño JA, Torres-Gonzalez F, Švab I, Rifel J, et al. An international risk prediction algorithm for the onset of generalized anxiety and panic syndromes in general practice attendees: predictA. *Psychol Med.* (2011) 41:1625–39. doi: 10.1017/S0033291710002400

22. Blanco C, Rubio J, Wall M, Wang S, Jiu CJ, Kendler KS. Risk factors for anxiety disorders: common and specific effects in a national sample. *Depress Anxiety.* (2014) 31:756–64. doi: 10.1002/da.22247

23. Druss BG, Rosenheck RA, Sledge WH. Health and disability costs of depressive illness in a major U.S. *corporation. Am J Psychiatry.* (2000) 157:1274–8. doi: 10.1176/appi.ajp.157.8.1274

24. Johnson J, Weissman MM, Klerman GL. Service utilization and social morbidity associated with depressive symptoms in the community. *JAMA.* (1992) 267:1478–83. doi: 10.1001/jama.267.11.1478

25. Berk M, Sarris J, Coulson CE, Jacka FN. Lifestyle management of unipolar depression. *Acta Psychiatr Scand.* (2013) 127:38–54. doi: 10.1111/acps.12124

26. Sarris J, O'Neil A, Coulson CE, Schweitzer I, Berk M. Lifestyle medicine for depression. *BMC Psychiatry.* (2014) 14:1–13. doi: 10.1186/1471-244X-14-107

27. Moylan S, Jacka FN, Pasco JA, Berk M. How cigarette smoking may increase the risk of anxiety symptoms and anxiety disorders: a critical review of biological pathways. *Brain Behav.* (2013) 3:302–26. doi: 10.1002/brb3.137

28. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. *Global Health.* (2020) 16:57. doi: 10.1186/s12992-020-00589-w

29. Choi EPH, Hui BPH, Wan EYF. Depression and anxiety in Hong Kong during COVID-19. *Int J Environ Res Public Health.* (2020) 17:3740. doi: 10.3390/ijerph17103740

30. Shevlin M, McBride O, Murphy J, Miller JG, Hartman TK, Levita L, et al. Anxiety, depression, traumatic stress, and COVID-19-related anxiety in the UK general population during the COVID-19 pandemic. *BJPsych Open.* (2020) 6:e125. doi: 10.1192/bjo.2020.109

31. Rezaee M, Putrenko I, Takeh A, Ganna A, Ingelsson E. Development and validation of risk prediction models for multiple cardiovascular diseases and Type 2 diabetes. *PLoS ONE.* (2020) 15:e0235758. doi: 10.1371/journal.pone.0235758

32. Alaa AM, Bolton T, Di Angelantonio E, Rudd JHF, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. *PLoS ONE.* (2019) 14:e0213653. doi: 10.1371/journal.pone.0213653

33. Köhler CA, Evangelou E, Stubbs B, Solmi M. Mapping risk factors for depression across the lifespan: an umbrella review of evidence from meta-analyses and Mendelian randomization studies. *J Psychiatr.* (2018) 103:189–207. doi: 10.1016/j.jpsychires.2018.05.020

34. Stegenga BT, King M, Grobbee DE, Torres-González F, Švab I, Maaroos H-I, et al. Differential impact of risk factors for women and men on the risk of major depressive disorder. *Ann Epidemiol.* (2012) 22:388–96. doi: 10.1016/j.annepidem.2012.04.011

35. Davidson-Pilon C, Kalderstam J, Jacobson N, Reed S, Kuhn B, Zivich P, et al. *CamDavidsonPilon/Lifelines: v0.25.6.* (2020). Available online at: https://zenodo.org/record/4457577 (accessed February 15, 2021).

36. Kvamme H, Borgan Ø, Scheel I. Time-to-event prediction with neural networks and cox regression. *J Mach Learn Res.* (2019) 20:1–30.

37. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, d/textquotesingle Alché-Buc F, Fox E, Garnett R, editors. *Advances in Neural Information Processing Systems 32.* Red Hook, NY: Curran Associates, Inc. (2019). p. 8024–35. Available online at: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

38. Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization. In: *25th Annual Conference on Neural Information Processing Systems (NIPS 2011).* Neural Information Processing Systems Foundation. (2011) Available online at : https://hal.inria.fr/hal-00642998/

39. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: a Next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* New York, NY: Association for Computing Machinery (2019). p. 2623–31. doi: 10.1145/3292500.3330701

40. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature.* (1986) 323:533–6. doi: 10.1038/323533a0

41. Kingma DP, Ba J. Adam: a method for stochastic optimization. *arXiv:14126980.* (2014).

42. Sarris J, Thomson R, Hargraves F, Eaton M, de Manincor M, Veronese N, et al. Multiple lifestyle factors and depressed mood: a cross-sectional and longitudinal analysis of the UK Biobank (N= 84,860). *BMC Med.* (2020) 18:1–10. doi: 10.1186/s12916-020-01813-5

43. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK biobank participants with those of the general population. *Am J Epidemiol.* (2017) 186:1026–34. doi: 10.1093/aje/kwx246

44. Bailey RK, Mokonogho J, Kumar A. Racial and ethnic differences in depression: current perspectives. *Neuropsychiatr Dis Treat.* (2019) 15:603–9. doi: 10.2147/NDT.S128584

45. Bellon JA, Moreno-Peral P, Moreno-Küstner B, Motrico E, Aiarzagüena JM, Fernández A, et al. Patients' opinions about knowing their risk for depression and what to do about it. The predictD-qualitative study. *PLoS ONE.* (2014) 9:e92008. doi: 10.1371/journal.pone.0092008

46. Wang J, Eccles H, Nannarone M, Schmitz N, Patten S, Lashewicz B. Does providing personalized depression risk information lead to increased psychological distress and functional impairment? Results from a mixed-methods randomized controlled trial. *Psychol Med.* (2020):1–9. doi: 10.1017/S0033291720003955

47. Pinto Pereira SM, Geoffroy M-C, Power C. Depressive symptoms and physical activity during 3 decades in adult life: bidirectional associations in a prospective cohort study. *JAMA Psychiatry.* (2014) 71:1373–80. doi: 10.1001/jamapsychiatry.2014.1240