Hierarchy of Demographic and Social Determinants of Mental Health

Jerzy Bala[1], PhD, Jennifer J. Newson[1], PhD, and Tara C. Thiagarajan[1]*, PhD.

[1]Sapien Labs, 1201 Wilson Blvd, 27th floor, Arlington, Virginia, United States of America

**** Corresponding Author***

Tara C. Thiagarajan
Sapien Labs
1201 Wilson Blvd, 27th floor
Arlington, VA
USA
Phone: 1 7039976657
Email: tara@sapienlabs.org

Coauthor Emails:

Jerzy Bala: jerzy@sapienlabs.org
Jennifer Newson: jennifer@sapienlabs.org

ABSTRACT

Objectives:

To understand the extent to which various demographic and social determinants predict mental wellbeing status and their relative hierarchy of predictive power in order to prioritize and develop population-based preventative approaches.

Design:

Cross-sectional analysis of survey data.

Setting:

Internet based survey from 32 countries across North America, Europe, Latin America, Middle East and North Africa, Sub Saharan Africa, South Asia and Australia.

Participants:

270,000 adults aged 18-85+ who participated in the Mental Health Million project.

Primary and secondary outcome measures:

We utilized 120+ demographic and social determinants to predict the aggregate mental health score of individuals (MHQ) and determine their relative predictive influence using various types of machine learning models including random forest, gradient boosting and logistic regression. The MHQ is derived from self-ratings of 47 mental health elements spanning ten disorders and provides a score that positions individuals along a spectrum from negative to positive mental health status that aligns with life impact and function criterion.

Results:

Classification models correctly identified 80% of those with a negative MHQ, while regression models predicted the specific MHQ score within ±15% of the position on the scale. Factors with the biggest predictive impact were young age followed by frequency of social interaction with friends, frequency of good sleep and physical exercise, and number of traumatic experiences. Age had twice the predictive power of social interaction which, in turn, was twice as important as the next four most important factors. Other predictive factors included sexual abuse, cyberbullying, and use of sleeping pills and sedatives.

ABSTRACT

Conclusion:

Social determinants of traumas and adversities and lifestyle can account for 60-70% of mental health challenges. However, additional factors are at play, particularly in younger age groups that are not included in this data and need further investigation.

Strengths and limitations of this study:

– The findings are based on a very large-scale global dataset that encompasses comprehensive mental health profiles and a wide array of demographics and social determinants.

– The MHQ provides an aggregate metric of mental health for overall prediction that has been validated against metrics of function such as work absenteeism and presenteeism as well as clinical diagnoses.

– The disproportionate impact of age indicates that important factors exist that have not been included heresuch as social factors (e.g. Internet behavior), dietary factors and other factors of the physical environment that cannot be easily captured through survey.

– Data are based on online self-report and therefore relevant only to an internet enabled audience, which excludes the poorest populations of the world where different factors may be at play. This approach is also not likely to fully capture the negative extreme, i.e., those with very severe mental illness who are not capable of accurate online self-assessment.

Keywords:

mental health; psychiatry; public health; risk factors; predictive modelling; machine learning

1. INTRODUCTION

The increased prevalence of mental health conditions as a consequence of the Covid-19 pandemic [1] and changing societal and generational dynamics [2-5] is placing increasing pressure on healthcare services [6]. This has created an urgent need to better understand the differential impact of various demographic and social determinants on mental health status. Such understanding can inform targeted preventative public health strategies at a population level to enhance societal mental health outcomes.

A number of determinants have been shown individually to contribute to mental health outcomes including socioeconomic status [7-9], employment status [10-12], educational attainment [13-15], sexual abuse [16-18], cyberbullying [19-22], divorce [23-26], physical exercise [27-31], social interaction [32-37] and sleep quality [38-41]. However, studies to date have focused either on individual social determinants, individual mental health disorders or specific populations or clinical groups. Consequently, we presently lack an integrated understanding of the core determinants which are universally most influential to people's mental health status and their relative importance, across multiple determinants, mental health disorders and population groups [42-44]. This understanding will provide guidance on how resources and public health strategies and initiatives can be deployed at a population level for maximal impact, and contribute to the ongoing debate on the extent to which mental health challenges can be addressed through societal rather than medical means [45-47].

Supervised and unsupervised machine learning approaches using large-scale data offer considerable opportunities for the advancement of mental health care and research [42] [48-52], and have been increasingly used to understand how multiple factors come together to predict health outcomes and their relative importance. This approach has been utilized with success in other fields such as cardiology [53-55]. However, data that aggregates many social determinants into a single study across a large population are rare. Medical records typically do not contain information on social determinants and what is available tends to be unstructured and must be mined from physician notes [56] [57]. Further medical records exclude the well population and therefore the ability to understand those social determinants that separate those with challenges from the well. Another challenge in mental health is that assessments are generally at the level of particular disorders and therefore do not provide an outcome of overall mental distress that aggregates across symptoms and disorders that tend to have high comorbidity [58-63]. Thus, while these techniques have been utilized to understand the social determinants of health generally [51] [64-66], they have not, to our knowledge, been used to predict mental health status from a large number of demographic and social determinants.

In this study, we used a unique global cross-sectional sample of 270,000 records spanning 32 countries and four languages taken from the Mental Health Million Project, a dynamic repository of global mental health data [67]. This data is obtained through the online MHQ assessment that includes self-assessment of 47 different elements of mental health, covering ten mental health disorders, on a life-impact scale, as well as self-report of over 120 potential determinants including demographics, lifestyle, trauma and adversity, substance use and medical conditions 68. An aggregate score of mental health, the mental health quotient or MHQ, positions individuals on a spectrum from distressed to thriving, and decreases systematically with loss of work productivity and increasing number of clinical symptoms [69] [70]. The objective of this study was therefore to use supervised learning approach to identify how well these demographic and social determinants could predict mental health status, as captured by the MHQ, and reveal the relative hierarchy of influence across these determinants.

2. MATERIALS AND METHODS

2.1. Data source and structure

The data used in this cross-sectional study was from the Mental Health Million Project, a dynamic, ongoing repository of global mental health and life context data that is openly available to the research community [67] and is acquired through the online MHQ assessment. This free and anonymous assessment captures ratings of 47 mental health elements on a life impact scale spanning symptoms of ten major mental health disorders and elements from the Research Domain Criteria (RDoC), as well as numerous life context factors including demographics, lifestyle factors, trauma experiences, medical conditions and substance use [67] [69 70]. It takes approximately 15 minutes to complete and returns a detailed personalized report to respondents. No financial compensation was provided. Participants were recruited via outreach campaigns on Facebook and using Google Ads by targeting a broad cross section of internet-connected adults 18-85 in each age-gender group and across a wide geographic and socioeconomic demographics to ensure broad representation in the sample. This online recruitment method provided a rapid, flexible, low-cost and anonymized way of sampling a broad cross section of the general population aged 18+ and overcomes many of the obstacles of probability sampling [71]. The final sample included 284,000 respondents, aged 18+, who had completed the MHQ between April 2020 and December 2021. This sample population spanned 32 countries and four languages (English, Spanish, French, Arabic; see Supplementary Table 1 for full list of countries and N values), where the online MHQ assessment was active during this time period. Records were removed if time to completion was <7 minutes, if the same option was selected for all rating questions (standard deviation of answers <0.5), or if the respondent provided incorrect or impossible answers (e.g., 500 hours since last meal). 270,000 records were included in the final analysis. The data used in this study received ethics approval from the Health Media Lab Institutional Review Board (Office for Human Research Protections Institutional Review Board #00001211, Federal Wide Assurance #00001102, IORG #0000850).

2.2. The Mental Health Quotient (MHQ)

The Mental Health Quotient or MHQ is an aggregate score that positions individuals on a spectrum from distressed to thriving [70]. The score is based on an algorithm that thresholds ratings as negative and positive based on the impact to function and applies a nonlinear transformation of the scale such that increasing negative impact to function is amplified [70]. The resulting MHQ scores fall on a positive-negative continuum. The positive scores, indicating normal functioning, range from 0 to 200 and are scaled to a mean of 100 based on sample data from 2019 (obtained from USA, UK, India English speaking population pre-COVID-19 pandemic). The negative side of the scale has the structure of a long tail that has been linearly rescaled to compress values within a range of -1 to -100 (to mitigate the impact of negative scores on the individual). (See Supplementary Figure 1 for re-scaled and original distributions of this data).

The MHQ score has been shown to have strong sample-to-sample consistency as well as criterion validity using data from 179,298 people across eight English-speaking countries [69]. This includes demonstration that, in the aggregate, average number of clinical symptoms and clinical diagnoses increase systematically as MHQ scores decrease, and that MHQ scores are linearly related to work productivity, including absenteeism and presenteeism [69]. Population MHQ scores also align with well-established trends relating to age, employment, education, physical exercise, sleep and social engagement, as well as being generally higher in males than females [72].

2.3. Encoding of demographic and social determinants

The various demographics and social determinants captured are shown in Supplementary Tables 2 and 3. Only surveys where all questions had been completed are submitted to the Mental Health Million database, there is thus no missing data within the variables of interest. Household income and ethnicity were not used since they were only obtained for select countries. Furthermore, household income could not be easily normalized across countries due to differences in currencies and purchasing power parity.

Within the MHQ, these determinants could be represented by two categories of data: categorical or numerical.

For the supervised learning approaches described below, a multiple-choice encoding method was used where items in multiple-choice lists (e.g., different types of trauma experiences) were each considered as individual features coded as either one (if selected) or zero (if not selected). Overall, this coding resulted in a feature set of 121 elements.

2.4. Correlations between determinants

Given that demographic, social and life experience factors are known to be correlated, as a check of the data we assessed correlations between different determinants. This was done by converting questions with ordinal answer selections into their numerical rank order. For example, selections of a single answer from among four options describing frequency of physical exercise and ranging from "Rarely/never" to "Everyday" were assigned numeric values from one to four representing increasing exercise frequency (Supplementary Table 4). These numerical mappings were then used to determine Spearman's rank correlations and their corresponding p-values (Supplementary Figure 2). The strongest positive correlations ($p<0.0001$) were between age and number of medical conditions ($r = 0.3$); number of traumas and substance use ($r = 0.2$); and frequency of good sleep and social interactions ($r = 0.2$). Physical exercise was also significantly positively correlated with sleep and social interaction ($r = 0.1$). Conversely, frequency of good sleep and social interaction were both negatively correlated with number of traumas experienced and number of substances used ($r = -0.2$). These findings are consistent with reports in the literature and provide additional validity to the data [73] [74].

2.5. Prediction of MHQ scores

To determine how well contextual factors predicted mental health status, defined as having either a positive or negative MHQ score, the following supervised learning models were used: random forest [75], gradient-boosting [76], naïve bayes [77], and logistic regression [78]. Three-, five-, and ten-fold cross-validation was performed, and five performance metrics were used to evaluate the four algorithms [area under the ROC curve (AUC), classification accuracy (CA), F1 measure, precision, recall]. Results were reported as the average of the positive MHQ and negative MHQ prediction models. Lift scores were calculated as the ratio between the true positive rate of the model and the positive rate in the population [79].

Separately, gradient boosting, random forest, and linear regression models [80] were used to predict individual MHQ scores across the -100 to +200 score range. The prediction performance was evaluated using root mean squared error (RMSE), mean absolute error (MAE), and R-squared (coefficient of determination).

All analysis was carried out using Python (version 3.8) including the scikit-learn, pandas, seaborn and shap libraries. Orange (version 3.32), an open-source Python library with a hierarchically-organized toolbox of data mining components, was used to simplify data manipulation, transformation, visualization, and modeling workflows.

2.6. Assessing the impact of feature categories and individual features

To assess the impact of different categories of features on model performance, feature categories were sequentially added (sequential forward selection), and performance metrics were calculated for the gradient boosting classification and regression models used for sign and score prediction, respectively.

The relative importance of individual features for the gradient boosting models was determined by how often the feature was selected to split the tree during learning, and how much it contributed to reducing the squared error over all trees in the model. The reduction in squared error attributable to that feature was computed based on the difference in squared error between that node and its children and normalized to the highest value. Thus, the larger the difference in squared error between the node and its children across the tree, the greater the influence of the feature. We also used the SHAP method to compute Shapley values, to assess how specific features affect prediction outcomes through additive feature attribution, thereby providing a view of both the magnitude and direction of each feature's contribution [81].

2.7. Patient and public involvement statement

There was no patient or public involvement in this study.

3. RESULTS

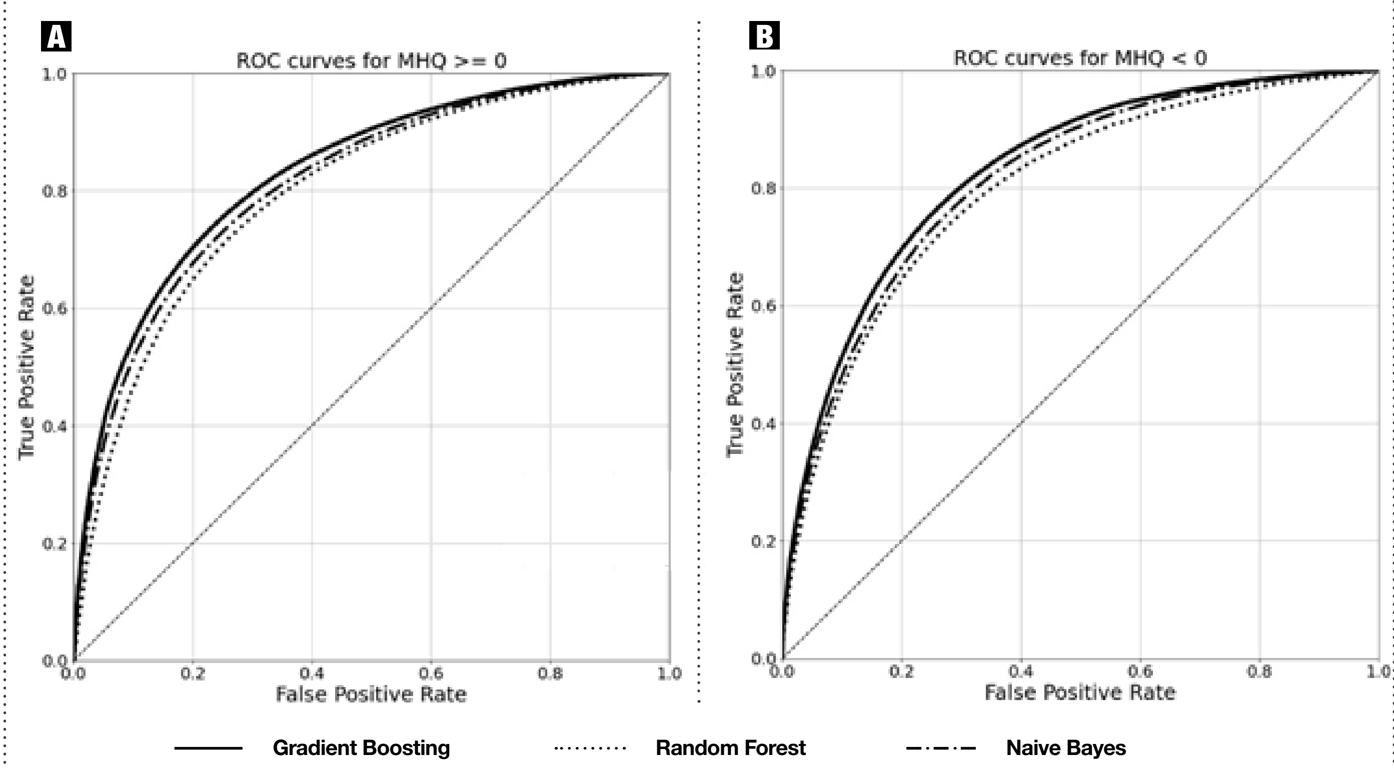
3.1. Prediction of mental health status and scores

All models (gradient boosting, random forest, naïve bayes and logistic regression) were able to classify mental health status as negative or positive using demographic and social determinants with high accuracy and precision (Supplementary Table 5). Negative status refers to MHQ scores <0 which typically represent ≥5 symptoms and functional impact of ≥3 days of work loss per month while positive refers to a normal range of function with typically <3 days of loss of work per month 69. ROC curves for model prediction of MHQ scores <0 (Figure 1A) and of MHQ scores ≥0 (Figure 1B) show that performance was similar for all models (AUC ranging from 0.8 for random forest to 0.84 for gradient boosting (GB); Supplementary Table 5). Going forward, we further characterized performance, robustness and feature importance for the GB model alone.

Overall, the GB classification model was able to correctly identify 80% of those who were struggling with their mental health (i.e., MHQ scores <0) with a precision of 79% (Supplementary Table 5). Across the range of scores, 85% of those with the most severe mental health challenges (lowest 5% of MHQ scores, typically corresponding to the presence of one or more clinical disorders 69 70 could be accurately identified as having negative mental health status (Supplementary Table 6). Conversely, 96% of those within the top 20% of MHQ scores (typically MHQ scores >120) could be correctly identified as having positive mental health status. The lift of the model (a measure of how much prediction is improved by the model relative to random classification) was >2 for the lowest 50% of negative scores and 2.7 for the lowest 5% (Supplementary Table 6).

Using GB, random forest and linear regression models, an individual's specific MHQ score could be predicted with an average error of ±18-19% of the 300-point MHQ scale using the RMSE methods and ±15-15.3% using the MAE method (Supplementary Table 7). Models had R2 between the actual and predicted MHQ values ranging from 0.4 for random forest to 0.44 for linear regression.

Figure 1 - *ROC CURVES FOR FOUR TYPES OF CLASSIFICATION MODELS PREDICTING* **(A)** *MHQ<0 AND* **(B)** *MHQ≥0.*



We next looked at how model performance changed with sample size for both the GB classification and regression models (Figure 2). Increasing sample size from 5,000 to 10,000 records provided the steepest gains in classification model performance and was relatively stable beyond a sample size of 50,000. For the regression model there were substantial performance gains as the sample size increased up to 40,000. Although not shown, we note that the standard deviation of performance metrics across iterations also decreased sharply as the sample size increased. This makes the case for the need for large scale studies of at least 20,000 to 50,000 for stable and robust results.

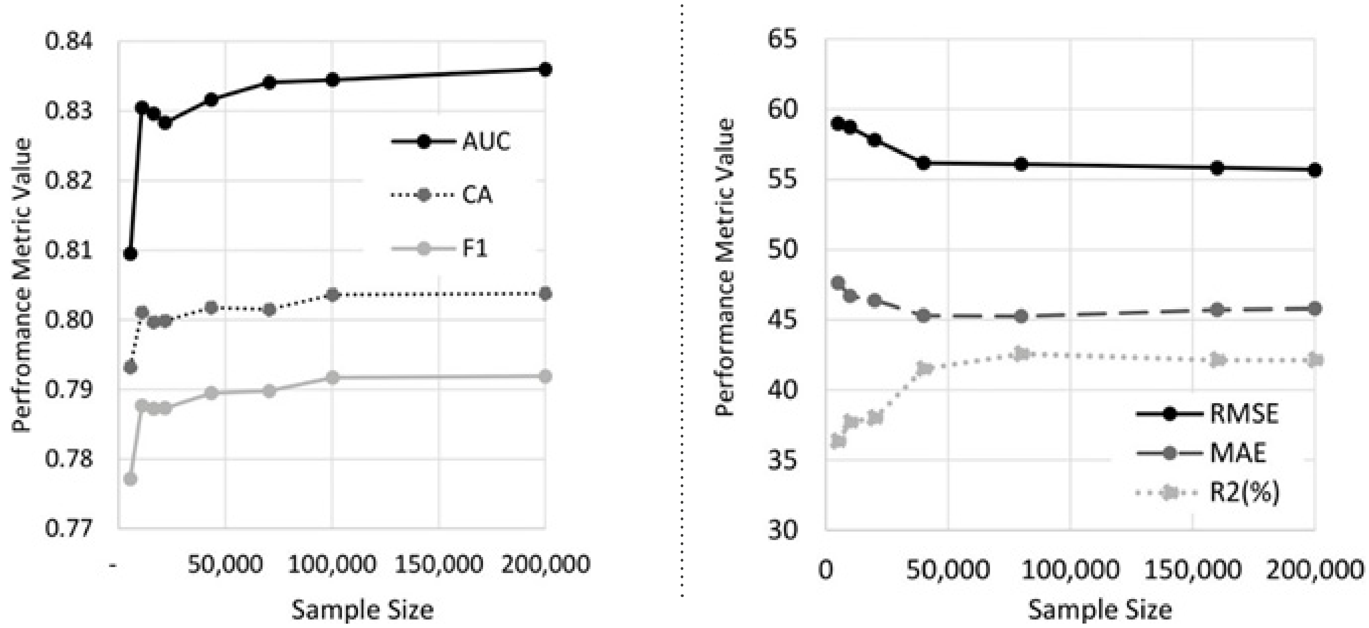
3.2. Contribution and relative importance of categories of determinants to prediction of mental health status

Evaluation of the GB classification model performance for different categories of determinants (Supplementary Table 8) showed that a subset of demographic features alone (age, gender, country, and language) predicted the sign of the MHQ score with an AUC of 0.75 and F1 of 0.73. Incorporating education attainment and employment status into the model increased the AUC and F1 scores to 0.78 and 0.75, respectively. Similarly, lifestyle factors alone (frequency of getting a good night's sleep; frequency of exercise; frequency of in person socializing with friends) had an AUC of 0.73 and an F1 score of 0.71, while trauma and adversities alone had a slightly lower AUC of 0.64 and an F1 score of 0.68. Combining all demographic factors with all lifestyle factors increased the AUC from 0.77 to 0.82 and F1 scores from 0.74 to 0.78. The further addition of traumas and adversities and medical conditions yielded no additional model improvement, while the addition of substance use marginally improved performance to an AUC of 0.84 and an F1 score of 0.79. A similar pattern of contribution of these determinant categories to model performance was observed for prediction of specific MHQ scores using regression (not shown). This redundancy of determinant categories reflects the interdependence of determinants and suggests that lifestyle and trauma experiences may derive in large part from demographic position.

Figure 2 - *PERFORMANCE CHARACTERISTICS OF THE GRADIENT BOOSTING MODELS*

A | Classification model performance characteristics increase with data sample size.

B | Regression model performance characteristics increase with data sample size.



3.3. Relative importance of specific individual determinants in the prediction of mental health status

For both the GB classification and regression models, we evaluated the reduction in the squared error attributable to each individual factor to determine their relative importance in predicting mental health status (Figure 3). This provided an estimate of how much each determinant contributed across all different trees of the model, each representing a different constellation of factors overall. Across both types of models, the most important factor for predicting MHQ sign or score was being in an 18-24 age range, which contributed twice as much predictive power compared to the next most important factor, which was rarely or never socializing with friends in person. This was followed by being in an 25-34 age range, rarely getting a good night's sleep, rarely engaging in physical exercise, and a higher number of lifetime traumas and adversities, all of which contributed only 30-45% of the predictive power of rarely socializing with friends in person. Employment status also featured among the top 20. Among the various traumas and adversities, sexual abuse or assault and cyberbullying contributed most, while use of sedatives or sleeping pills contributed the most of all substances used. Notably, the experience of financial adversities were not individually prominent in prediction.

We also assessed whether a particular feature predicted a more negative or positive MHQ using SHAP values (Figure 4). The dominant factors were consistent with findings from the squared error method. Age under 35, lack of in-person socializing, poor sleep, lack of physical exercise, the experience of a larger number of traumas and adversities, sexual abuse, cyberbullying and use of sedatives or sleeping pills contributed strongly to negative or low MHQ scores, while regular in-person socializing, exercising, getting a good night's sleep, and older age contributed to positive MHQ scores.

Figure 3 - *TOP 20 FEATURES RANKED IN ORDER OF IMPORTANCE FOR MODEL PERFORMANCE FOR* **(A)** *GRADIENT BOOSTING CLASSIFICATION AND* **(B)** *GRADIENT BOOSTING REGRESSION. FEATURE IMPORTANCE IS COMPUTED AS THE SUM OF THE CONTRIBUTIONS TO SQUARED ERROR REDUCTION AND NORMALIZED TO THE VALUE FOR AGE*

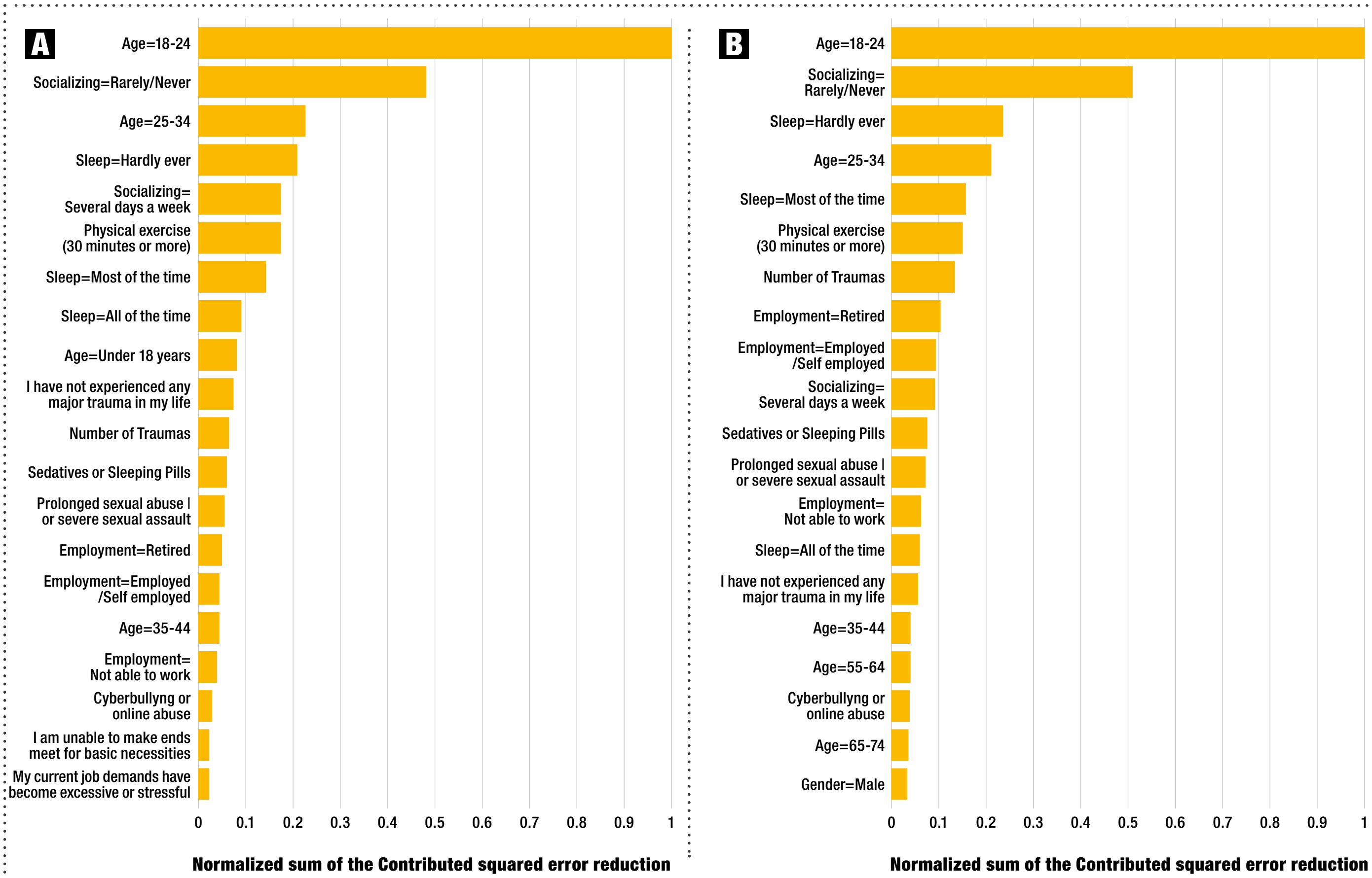
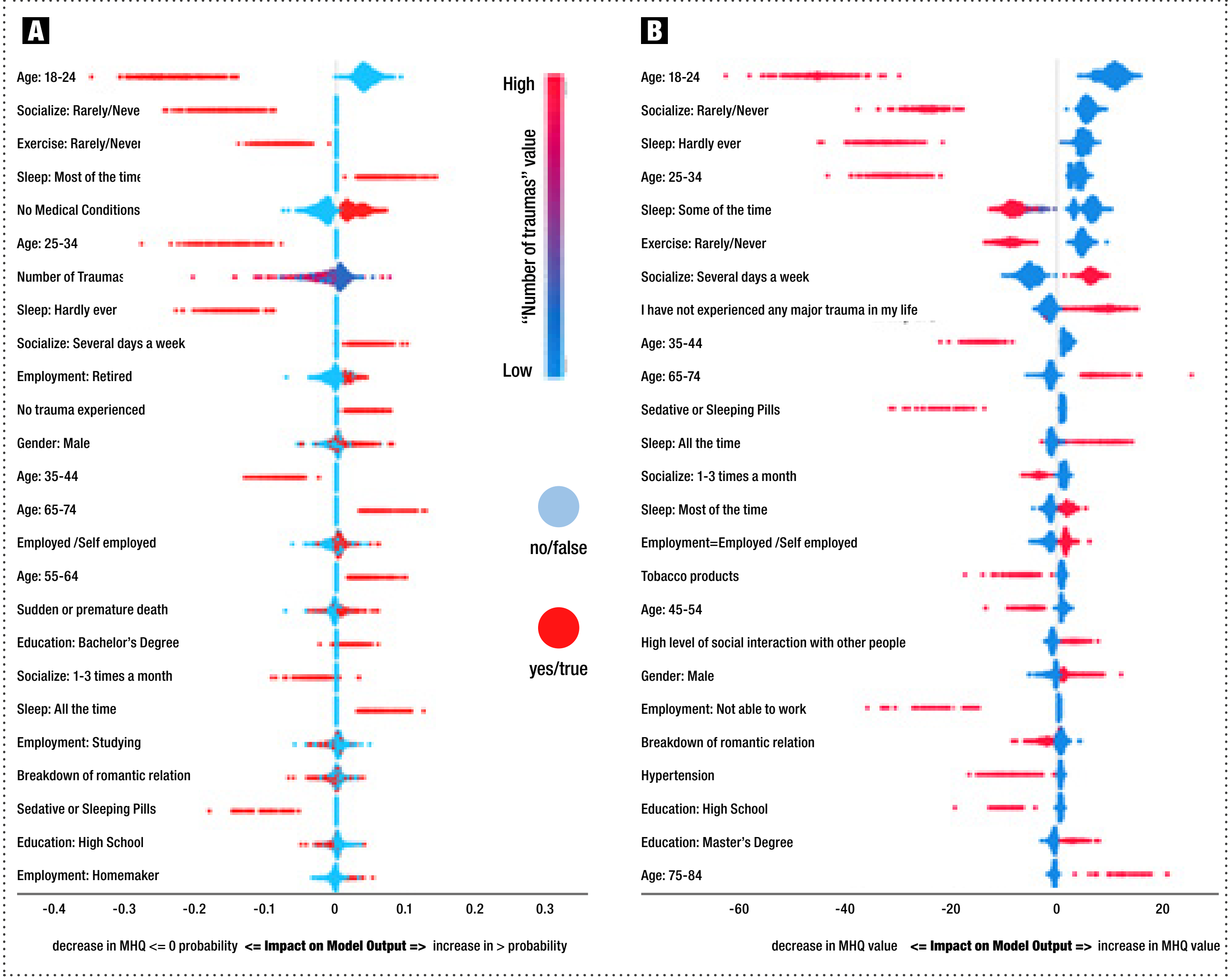


Figure 4 - *SHAP VALUES FOR THE TOP 25 FACTORS FOR* **(A)** *THE GRADIENT BOOSTING CLASSIFICATION MODEL AND* **(B)** *THE GRADIENT BOOSTING REGRESSION MODEL*



4. DISCUSSION

Using a comprehensive set of 47 mental health symptoms and 120+ determinants from 270,000 adults, we have shown, as per our objectives outlined in the introduction, that with just a handful of demographic and social determinants, it is possible to predict mental health status with high accuracy and precision. Furthermore, we show the hierarchy of importance of individual determinants and highlight the dominance of being part of GenZ (18-24) and infrequency of in-person socializing with friends as outsized predictors of mental health status.

4.1. Life context as the primary determinant of mental health status

We have shown here that 80% of people struggling with mental health challenges could be accurately identified from their demographic and social characteristics. Similarly, one's specific MHQ score or position on a scale of mental health ranging from negative to positive could be predicted within an average error of ±15%. This suggests that our mental health status is largely dependent on the societal milieu in which we live; in essence an expected response of our brain and mind to ongoing life circumstances [42] [44] [82-85]. Understanding these demographic and social determinants provides an opportunity to substantially alter mental health outcomes at a population level through systemic societal shifts and delivers an impetus for individuals to take action to alter the circumstances of their own lives.

4.2. The relative impact of individual determinants

This study furthers our understanding of the specific demographic and social determinants that are most influential in driving population mental health status. While all demographic factors together were effective at predicting MHQ sign, young age (i.e., being 18-24 followed by 25-34) was disproportionately powerful as a predictor of negative mental health. This is supported by other evidence that shows overall mental health status is worse for each younger age group: data from the Mental Health Million Project has shown that in 2021, 44% of young adults (18-24 years) were mentally distressed or struggling compared to 7% among those aged 65 and above [72], while other studies also highlight the increase in mental health problems in teens [2] [3]. This is in sharp contrast to psychological wellbeing patterns observed prior to 2010 where young adults were typically at the higher end of wellbeing scales [86] [87]. The timeline of this decline of younger generations is also highlighted by a recent CDC report that shows a sharp rise in feelings of sadness reported by teens only in this last decade [88].

Given that age is immutable, this trend suggests that age stands as a proxy for global changes in the environment and life context with each generation that are not currently captured in this data. Two key factors stand out as important for further investigation. The first is the considerable shift in the socio-technological environment across generations with the introduction of the internet in the 1990s and smart phones (e.g., iPhones) in 2007. A growing body of evidence suggests that this shift, and in particular the unhealthy use of social media, is having a negative effect on mental health within the GenZ population [89]. The second is the concerning many-fold increase in the levels of neuroendocrine disruptors and neurotoxic substances such as microplastics and phthalates in our food and water [90-92] that are increasingly detectable in our blood [93-95]. A corollary of this is that the impact of age on model outcomes will change with changing environment.

Among lifestyle or life experience factors, lack of in-person socializing with friends, lack of physical exercise, poor sleep and a larger number of traumatic experiences were key predictors of negative mental health. Interestingly, lack of in-person socializing with friends was almost twice as important as all the other factors and is supported by other evidence highlighting the importance of in person socializing [33] [96]. The reasons behind low levels of in-person socializing are complex, calling for the need to evaluate deeply the sociological factors that drive it. It is also of interest that cyberbullying, which is far more prevalent among younger adults [97], was one of the key trauma factors in addition to sexual abuse, and aligns with other research [16-22]. On the other hand, the experience of financial traumas and adversities such as homelessness and difficulty making ends meet were relatively less predictive and did not make it into the top 25. Together, this hierarchy of influence across demographic and social determinants provides an initial framework to approach population mental health at a preventative level and points to where efforts should be focused for the greatest impact. Solutions that enable greater frequency of in-person social interaction for instance could have a much greater impact on population mental health compared to financial programs while a combination of increased in-person social interaction and physical exercise may enhance population mental health more significantly than tackling the prevalence of a host of traumas and adversities.

4.3. Implications, limitations and conclusions

The degree to which demographic and social determinants predict our mental health status attests to how intimately intertwined the struggles of mind are with life circumstances. Practically, these findings are a first view of how analysis of large-scale global multidimensional cross-disorder data can provide insights into the relative impact of various demographic and social determinants on overall population mental health. An implication of understanding the full impact of these determinants is that it can enable the separation of mental health profiles that are predominantly socially driven from those that are predominantly biologically driven (i.e., due to genetics, pathogens, toxins). This first iteration, however, identifies certain gaps. First, the disproportionate impact of age indicates that important factors exist that have not been included here. The most obvious and significant relates to use of the internet. In addition, the importance of in-person socializing suggests that it is important to probe social relationships in much more detail. The addition of relevant factors in these areas may substantially elevate model performance and eliminate the requirement for age in the model, which likely changes in its predictive nature as environment and life context changes. This would then provide a clearer picture of social determinants to facilitate the design of effective interventions and policies. We also acknowledge that this data is based on online self-report and is therefore relevant only to an internet enabled audience, which excludes the poorest populations of the world where different factors may be at play.

A further limitation of this data is that it is not likely to fully capture the negative extreme, i.e., those with very severe mental illness who are not capable of accurate online self-assessment. However, while this approach may have these limitations, it provides a substantial view of the drivers of population mental health and adds a perspective to the debate on the extent to which mental health challenges can be addressed through societal rather than medical means [45-47].

In summary, we provide an initial view of the aggregate impact of demographic and social determinants on mental health status, and a hierarchy of determinants that can inform and enhance our ability to impact population mental health.

DATA AVAILABILITY STATEMENT

Data is available via the online data repository Brainbase where data from the ongoing Mental Health Million Project is dynamically updated as respondents complete the MHQ. The anonymized dataset, together with supporting information, is freely available for use for not-for profit purposes on request. See [67] for more information on how to request access.

FUNDING STATEMENT

This work was supported by funding from Sapien Labs. It did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

DECLARATIONS OF INTEREST

One author (TT) received a grant award from the National Institute of Mental Health (NIMH) to develop a commercial version of the MHQ tool referenced herein for clinical use. There are no other commercial or financial relationships.

ETHICS APPROVAL

The data used in this study received ethics approval from the Health Media Lab Institutional Review Board (Office for Human Research Protections Institutional Review Board #00001211, Federal Wide Assurance #00001102, IORG #0000850).

AUTHOR CONTRIBUTIONS

TT and JB were responsible for the study concept, design and interpretation. JB was responsible for the analysis. TT, JB and JN drafted the manuscript. All the authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

REFERENCES

1. WHO. Mental Health and COVID-19: Early evidence of the pandemic's impact: Scientific Brief Geneva: WHO; 2022 [Available from: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1.

2. Krokstad S, Weiss DA, Krokstad MA, et al. Divergent decennial trends in mental health according to age reveal poorer mental health for young people: repeated cross-sectional population-based surveys from the HUNT Study, Norway. *BMJ Open* 2022;12(5):e057654. doi: 10.1136/bmjopen-2021-057654

3. Twenge J, Cooper A, Joiner T, et al. Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005-2017. *J Abnorm Psychol* 2019;128(3):185-99. doi: 10.1037/abn0000410 [published Online First: 2019/03/15]

4. Bolton D, Bhugra D. Changes in society and young people's mental health1. *International Review of Psychiatry* 2021;33(1-2):154-61. doi: 10.1080/09540261.2020.1753968

5. Newson J, Pastukh V, Sukhoi O, et al. Mental State of the World 2020, Mental Health Million project. In: Labs S, ed., 2021.

6. Comission CQ. Monitoring the Mental Health Act in 2020/21 2020 [Available from: https://www.cqc.org.uk/publications/major-report/monitoring-mental-health-act.

7. Miech RA, Caspi A, Moffitt TE, et al. Low socioeconomic status and mental disorders: A longitudinal study of selection and causation during young adulthood. *American Journal of Sociology* 1999;104:1096-131. doi: 10.1086/210137

8. Kivimäki M, Batty GD, Pentti J, et al. Association between socioeconomic status and the development of mental and physical health conditions in adulthood: a multi-cohort study. *Lancet Public Health* 2020;5(3):e140-e49. doi: 10.1016/s2468-2667(19)30248-8 [published Online First: 2020/02/03]

9. Barbalat G, Franck N. Ecological study of the association between mental illness with human development, income inequalities and unemployment across OECD countries. *BMJ Open* 2020;10(4):e035055. doi: 10.1136/bmjopen-2019-035055

10. Bartelink VHM, Zay Ya K, Guldbrandsson K, et al. Unemployment among young people and mental health: A systematic review. *Scand J Public Health* 2020;48(5):544-58. doi: 10.1177/1403494819852847 [published Online First: 2019/07/12]

11. Mousteri V, Daly M, Delaney L, et al. Adolescent mental health and unemployment over the lifespan: Population evidence from Sweden. *Soc Sci Med* 2019;222:305-14. doi: 10.1016/j.socscimed.2018.12.030 [published Online First: 2019/01/25]

12. Mousteri V, Daly M, Delaney L. The scarring effect of unemployment on psychological well-being across Europe. *Soc Sci Res* 2018;72:146-69. doi: 10.1016/j.ssresearch.2018.01.007 [published Online First: 2018/04/04]

13. Cutler DM, Lleras-Muney A. Education and Health: Insights from International Comparisons. 2012 doi: 10.3386/w17738

14. Cohen AK, Syme SL. Education: a missed opportunity for public health intervention. *Am J Public Health* 2013;103(6):997-1001. doi: 10.2105/ajph.2012.300993 [published Online First: 2013/04/20]

15. Cohen AK, Nussbaum J, Weintraub MLR, et al. Association of Adult Depression With Educational Attainment, Aspirations, and Expectations. *Prev Chronic Dis* 2020;17:E94. doi: 10.5888/pcd17.200098 [published Online First: 2020/08/29]

16. Hailes HP, Yu R, Danese A, et al. Long-term outcomes of childhood sexual abuse: an umbrella review. *Lancet Psychiatry* 2019;6(10):830-39. doi: 10.1016/s2215-0366(19)30286-x [published Online First: 2019/09/15]

17. Maniglio R. The impact of child sexual abuse on health: a systematic review of reviews. *Clin Psychol Rev* 2009;29(7):647-57. doi: 10.1016/j.cpr.2009.08.003 [published Online First: 2009/09/08]

18. Chen LP, Murad MH, Paras ML, et al. Sexual abuse and lifetime diagnosis of psychiatric disorders: systematic review and meta-analysis. *Mayo Clin Proc* 2010;85(7):618-29. doi: 10.4065/mcp.2009.0583 [published Online First: 2010/05/12]

19. Giumetti GW, Kowalski RM. Cyberbullying via social media and well-being. *Curr Opin Psychol* 2022;45:101314. doi: 10.1016/j.copsyc.2022.101314 [published Online First: 2022/03/22]

20. Fahy AE, Stansfeld SA, Smuk M, et al. Longitudinal Associations Between Cyberbullying Involvement and Adolescent Mental Health. *J Adolesc Health* 2016;59(5):502-09. doi: 10.1016/j.jadohealth.2016.06.006 [published Online First: 2016/10/25]

21. van Geel M, Vedder P, Tanilon J. Relationship between peer victimization, cyberbullying, and suicide in children and adolescents: a meta-analysis. *JAMA Pediatr* 2014;168(5):435-42. doi: 10.1001/jamapediatrics.2013.4143 [published Online First: 2014/03/13]

22. Kowalski RM, Giumetti GW, Schroeder AN, et al. Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth. *Psychol Bull* 2014;140(4):1073-137. doi: 10.1037/a0035618 [published Online First: 2014/02/12]

23. Whisman MA, Salinger JM, Sbarra DA. Relationship dissolution and psychopathology. *Curr Opin Psychol* 2022;43:199-204. doi: 10.1016/j.copsyc.2021.07.016 [published Online First: 2021/08/21]

24. D'Onofrio B, Emery R. Parental divorce or separation and children's mental health. *World Psychiatry* 2019;18(1):100-01. doi: 10.1002/wps.20590 [published Online First: 2019/01/03]

25. Tullius JM, De Kroon MLA, Almansa J, et al. Adolescents' mental health problems increase after parental divorce, not before, and persist until adulthood: a longitudinal TRAILS study. *Eur Child Adolesc Psychiatry* 2022;31(6):969-78. doi: 10.1007/s00787-020-01715-0 [published Online First: 2021/02/11]

26. Auersperg F, Vlasak T, Ponocny I, et al. Long-term effects of parental divorce on mental health - A meta-analysis. *J Psychiatr Res* 2019;119:107-15. doi: 10.1016/j.jpsychires.2019.09.011 [published Online First: 2019/10/18]

27. Chekroud SR, Gueorguieva R, Zheutlin AB, et al. Association between physical exercise and mental health in 1•2 million individuals in the USA between 2011 and 2015: a cross-sectional study. *Lancet Psychiatry* 2018;5(9):739-46. doi: 10.1016/s2215-0366(18)30227-x [published Online First: 2018/08/14]

28. Cooney G. Exercise and mental health: a complex and challenging relationship. *Lancet Psychiatry* 2018;5(9):692-93. doi: 10.1016/s2215-0366(18)30291-8 [published Online First: 2018/08/14]

29. Schuch FB, Vancampfort D, Firth J, et al. Physical Activity and Incident Depression: A Meta-Analysis of Prospective Cohort Studies. *Am J Psychiatry* 2018;175(7):631-48. doi: 10.1176/appi.ajp.2018.17111194 [published Online First: 2018/04/25]

30. Pearce M, Garcia L, Abbas A, et al. Association Between Physical Activity and Risk of Depression: A Systematic Review and Meta-analysis. *JAMA Psychiatry* 2022;79(6):550-59. doi: 10.1001/jamapsychiatry.2022.0609 [published Online First: 2022/04/14]

31. McKeon G, Curtis J, Rosenbaum S. Promoting physical activity for mental health: an updated evidence review and practical guide. *Curr Opin Psychiatry* 2022;35(4):270-76. doi: 10.1097/yco.0000000000000796 [published Online First: 2022/06/09]

32. Morina N, Kip A, Hoppen TH, et al. Potential impact of physical distancing on physical and mental health: a rapid narrative umbrella review of meta-analyses on the link between social connection and health. *BMJ Open* 2021;11(3):e042335. doi: 10.1136/bmjopen-2020-042335 [published Online First: 2021/03/20]

33. Umberson D, Montez J. Social relationships and health: a flashpoint for health policy. *J Health Soc Behav* 2010;51 Suppl(Suppl):S54-66. doi: 10.1177/0022146510383501 [published Online First: 2010/12/22]

34. Santini ZI, Koyanagi A, Tyrovolas S, et al. The association between social relationships and depression: a systematic review. *J Affect Disord* 2015;175:53-65. doi: 10.1016/j.jad.2014.12.049 [published Online First: 2015/01/17]

35. Leigh-Hunt N, Bagguley D, Bash K, et al. An overview of systematic reviews on the public health consequences of social isolation and loneliness. *Public Health* 2017;152:157-71. doi: 10.1016/j.puhe.2017.07.035 [published Online First: 2017/09/16]

36. Sun J, Harris K, Vazire S. Is well-being associated with the quantity and quality of social interactions? *J Pers Soc Psychol* 2020;119(6):1478-96. doi: 10.1037/pspp0000272 [published Online First: 2019/10/28]

37. Kuczynski AM, Halvorson MA, Slater LR, et al. The effect of social interaction quantity and quality on depressed mood and loneliness: A daily diary study. *Journal of Social and Personal Relationships* 2021;39(3):734-56. doi: 10.1177/02654075211045717

38. Hill Almeida LM, Flicker L, Hankey GJ, et al. Disrupted sleep and risk of depression in later life: A prospective cohort study with extended follow up and a systematic review and meta-analysis. *J Affect Disord* 2022;309:314-23. doi: 10.1016/j.jad.2022.04.133 [published Online First: 2022/05/02]

39. Scott AJ, Webb TL, Martyn-St James M, et al. Improving sleep quality leads to better mental health: A meta-analysis of randomised controlled trials. *Sleep Med Rev* 2021;60:101556. doi: 10.1016/j.smrv.2021.101556 [published Online First: 2021/10/05]

40. Plante DT. The Evolving Nexus of Sleep and Depression. *Am J Psychiatry* 2021;178(10):896-902. doi: 10.1176/appi.ajp.2021.21080821 [published Online First: 2021/10/02]

41. Freeman D, Sheaves B, Waite F, et al. Sleep disturbance and psychiatric disorders. *Lancet Psychiatry* 2020;7(7):628-37. doi: 10.1016/s2215-0366(20)30136-x [published Online First: 2020/06/22]

42. Alegría M, NeMoyer A, Falgàs Bagué I, et al. Social Determinants of Mental Health: Where We Are and Where We Need to Go. *Curr Psychiatry Rep* 2018;20(11):95. doi: 10.1007/s11920-018-0969-9 [published Online First: 2018/09/18]

43. Shim RS, Compton MT. Addressing the Social Determinants of Mental Health: If Not Now, When? If Not Us, Who? *Psychiatr Serv* 2018;69(8):844-46. doi: 10.1176/appi.ps.201800060 [published Online First: 2018/06/02]

44. Handerer F, Kinderman P, Shafti M, et al. A Scoping Review and Narrative Synthesis Comparing the Constructs of Social Determinants of Health and Social Determinants of Mental Health: Matryoshka or Two Independent Constructs? *Front Psychiatry* 2022;13:848556. doi: 10.3389/fpsyt.2022.848556 [published Online First: 2022/05/03]

45. Editors PM. The paradox of mental health: over-treatment and under-recognition. *PLoS Med* 2013;10(5):e1001456. doi: 10.1371/journal.pmed.1001456 [published Online First: 2013/06/01]

46. Thangadurai P, Jacob KS. Medicalizing distress, ignoring public health strategies. *Indian J Psychol Med* 2014;36(4):351-4. doi: 10.4103/0253-7176.140698 [published Online First: 2014/10/23]

47. Clark J. Medicalization of global health 2: The medicalization of global mental health. *Glob Health Action* 2014;7:24000. doi: 10.3402/gha.v7.24000 [published Online First: 2014/05/23]

48. Koutsouleris N, Hauser TU, Skvortsova V, et al. From promise to practice: towards the realisation of AI-informed mental health care. *The Lancet Digital Health* 2022;4(11):e829-e40. doi: 10.1016/S2589-7500(22)00153-4

49. Deferio JJ, Breitinger S, Khullar D, et al. Social determinants of health in mental health care and research: a case for greater inclusion. *Journal of the American Medical Informatics Association* 2019;26(8-9):895-99. doi: 10.1093/jamia/ocz049

50. Osawa I, Goto T, Tabuchi T, et al. Machine-learning approaches to identify determining factors of happiness during the COVID-19 pandemic: retrospective cohort study. *BMJ Open* 2022;12(12):e054862. doi: 10.1136/bmjopen-2021-054862

51. Morgenstern JD, Buajitti E, O'Neill M, et al. Predicting population health with machine learning: a scoping review. *BMJ Open* 2020;10(10):e037860. doi: 10.1136/bmjopen-2020-037860

52. Dreger S, Buck C, Bolte G. Material, psychosocial and sociodemographic determinants are associated with positive mental health in Europe: a cross-sectional study. *BMJ Open* 2014;4(5):e005095. doi: 10.1136/bmjopen-2014-005095

53. Zhao Y, Wood EP, Mirin N, et al. Social Determinants in Machine Learning Cardiovascular Disease Prediction Models: A Systematic Review. *American Journal of Preventive Medicine* 2021;61(4):596-605. doi: https://doi.org/10.1016/j.amepre.2021.04.016

54. Segar MW, Hall JL, Jhund PS, et al. Machine Learning–Based Models Incorporating Social Determinants of Health vs Traditional Models for Predicting In-Hospital Mortality in Patients With Heart Failure. *JAMA Cardiology* 2022;7(8):844-54. doi: 10.1001/jamacardio.2022.1900

55. Yancy CW, Khan SS. Replacing Race With Social Determinants of Health in Risk Prediction—Getting It Right. *JAMA Cardiology* 2022;7(8):856-56. doi: 10.1001/jamacardio.2022.1909

56. Han S, Zhang RF, Shi L, et al. Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing. *Journal of Biomedical Informatics* 2022;127:103984. doi: https://doi.org/10.1016/j.jbi.2021.103984

57. Patra BG, Sharma MM, Vekaria V, et al. Extracting social determinants of health from electronic health records using natural language processing: a systematic review. *J Am Med Inform Assoc* 2021;28(12):2716-27. doi: 10.1093/jamia/ocab170 [published Online First: 2021/10/07]

58. Newson JJ, Hunter D, Thiagarajan TC. The Heterogeneity of Mental Health Assessment. *Front Psychiatry* 2020;11:76. doi: 10.3389/fpsyt.2020.00076 [published Online First: 2020/03/17]

59. Hirschfeld RMA. The Comorbidity of Major Depression and Anxiety Disorders: Recognition and Management in Primary Care. *Prim Care Companion J Clin Psychiatry* 2001;3(6):244-54. doi: 10.4088/pcc.v03n0609

60. Kessler RC, Merikangas KR, Wang PS. Prevalence, comorbidity, and service utilization for mood disorders in the United States at the beginning of the twenty-first century. *Annu Rev Clin Psychol* 2007;3:137-58. doi: 10.1146/annurev.clinpsy.3.022806.091444 [published Online First: 2007/08/25]

61. Brady KT. Posttraumatic stress disorder and comorbidity: recognizing the many faces of PTSD. *J Clin Psychiatry* 1997;58 Suppl 9:12-5. [published Online First: 1997/01/01]

62. Allsopp K, Read J, Corcoran R, et al. Heterogeneity in psychiatric diagnostic classification. *Psychiatry Research* 2019;279:15-22. doi: https://doi.org/10.1016/j.psychres.2019.07.005

63. Angold A, Costello EJ, Erkanli A. Comorbidity. *Journal of Child Psychology and Psychiatry* 1999;40(1):57-87. doi: https://doi.org/10.1111/1469-7610.00424

64. Burgermaster M, Rodriguez VA. Psychosocial-Behavioral Phenotyping: A Novel Precision Health Approach to Modeling Behavioral, Psychological, and Social Determinants of Health Using Machine Learning. *Ann Behav Med* 2022 doi: 10.1093/abm/kaac012 [published Online First: 2022/04/22]

65. Kino S, Hsu YT, Shiba K, et al. A scoping review on the use of machine learning in research on social determinants of health: Trends and research prospects. *SSM Popul Health* 2021;15:100836. doi: 10.1016/j.ssmph.2021.100836 [published Online First: 2021/06/26]

66. Seligman B, Tuljapurkar S, Rehkopf D. Machine learning approaches to the social determinants of health in the health and retirement study. *SSM - Population Health* 2018;4:95-99. doi: https://doi.org/10.1016/j.ssmph.2017.11.008

67. Newson J, Thiagarajan T. Dynamic Dataset of Global Population Mental Wellbeing. *PsyArXiv* 2021 doi: 10.31234/osf.io/vtzne

68. Newson J, Hunter D, Thiagarajan T. The Heterogeneity of Mental Health Assessment. *Front Psychiatry* 2020;11:76. doi: 10.3389/fpsyt.2020.00076 [published Online First: 2020/03/17]

69. Newson J, Pastukh V, Thiagarajan T. Assessment of Population Well-being With the Mental Health Quotient: Validation Study. *JMIR Mental Health* 2022;9(4):e34105. doi: 10.2196/34105

70. Newson J, Thiagarajan T. Assessment of Population Well-Being With the Mental Health Quotient (MHQ): Development and Usability Study. *JMIR Ment Health* 2020;7(7):e17935. doi: 10.2196/17935

71. Schneider D, Harknett K. What's to Like? Facebook as a Tool for Survey Data Collection. *Sociological Methods & Research* 2022;51(1):108-40. doi: 10.1177/0049124119882477

72. Newson J, Pastukh V, Sukhoi O, et al. *Mental State of the World Reports* 2022 [updated March. Available from: https://mentalstateoftheworld.report/.

73. Degenhardt L, Bharat C, Glantz M, et al. The associations between traumatic experiences and subsequent onset of a substance use disorder: Findings from the World Health Organization World Mental Health surveys. *Drug and Alcohol Dependence* 2022:109574. doi: https://doi.org/10.1016/j.drugalcdep.2022.109574

74. Brownlow JA, Klingaman EA, Miller KE, et al. Trauma type as a risk factor for insomnia in a military population. *J Affect Disord* 2022;308:65-70. doi: 10.1016/j.jad.2022.04.018 [published Online First: 2022/04/13]

75. Ho TK. Random decision forests. . *Proceedings of 3rd international conference on document analysis and recognition* 1995;1:278–82.

76. Friedman J, H. . Greedy function approximation: A gradient boosting machine. *The Annals of Statistics* 2001;29(5):1189-232. doi: 10.1214/aos/1013203451

77. Vikramkumar. Bayes and Naive Bayes Classifier. *ArXiv* 2014 doi: abs/1404.0933

78. Cox DR. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society Series B (Methodological)* 1958;20(2):215-42.

79. Piatetsky-Shapiro G, Masand B. Estimating campaign benefits and modeling lift. *KDD* 1999;99:185–93.

80. Yan X, Su X. Linear Regression Analysis: Theory and Computing: World Scientific 2009.

81. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA: Curran Associates Inc., 2017:4768–77.

82. Allen J, Balfour R, Bell R, et al. Social determinants of mental health. *Int Rev Psychiatry* 2014;26(4):392-407. doi: 10.3109/09540261.2014.928270 [published Online First: 2014/08/20]

83. Compton MT, Shim RS. The Social Determinants of Mental Health. *FOCUS* 2015;13(4):419-25. doi: 10.1176/appi.focus.20150017

84. İlgun G. What are the socioeconomic determinants of mental disorders? *Perspectives in Psychiatric Care* 2022;n/a(n/a) doi: https://doi.org/10.1111/ppc.13136

85. Silva M, Loureiro A, Cardoso G. Social determinants of mental health: A review of the evidence. *The European Journal of Psychiatry* 2016;30:259-92.

86. Steptoe A, Deaton A, Stone AA. Subjective wellbeing, health, and ageing. *Lancet* 2015;385(9968):640-48. doi: 10.1016/s0140-6736(13)61489-0 [published Online First: 2014/12/04]

87. Stone AA, Schwartz JE, Broderick JE, et al. A snapshot of the age distribution of psychological well-being in the United States. *Proceedings of the National Academy of Sciences* 2010;107(22):9985-90. doi: 10.1073/pnas.1003744107

88. CDC. Youth Risk Behavior Survey. *Data Summary & Trends Report* 2023 [Available from: https://www.cdc.gov/healthyyouth/data/yrbs/pdf/YRBS_Data-Summary-Trends_Report2023_508.pdf.

89. Twenge J. Increases in Depression, Self-Harm, and Suicide Among U.S. Adolescents After 2012 and Links to Technology Use: Possible Mechanisms. *Psychiatric Research and Clinical Practice* 2020;2(1):19-25. doi: https://doi.org/10.1176/appi.prcp.20190015

90. Vivekanand AC, Mohapatra S, Tyagi VK. Microplastics in aquatic environment: Challenges and perspectives. *Chemosphere* 2021;282:131151. doi: 10.1016/j.chemosphere.2021.131151 [published Online First: 2021/09/03]

91. Ragusa A, Svelato A, Santacroce C, et al. Plasticenta: First evidence of microplastics in human placenta. *Environ Int* 2021;146:106274. doi: 10.1016/j.envint.2020.106274 [published Online First: 2021/01/06]

92. Gani KM, Tyagi VK, Kazmi AA. Occurrence of phthalates in aquatic environment and their removal during wastewater treatment processes: a review. *Environ Sci Pollut Res Int* 2017;24(21):17267-84. doi: 10.1007/s11356-017-9182-3 [published Online First: 2017/06/02]

93. Grandjean P, Landrigan PJ. Neurobehavioural effects of developmental toxicity. *Lancet Neurol* 2014;13(3):330-8. doi: 10.1016/s1474-4422(13)70278-3 [published Online First: 2014/02/22]

94. Leslie HA, van Velzen MJM, Brandsma SH, et al. Discovery and quantification of plastic particle pollution in human blood. *Environment International* 2022;163:107199. doi: https://doi.org/10.1016/j.envint.2022.107199

95. Sicińska P, Kik K, Bukowska B. Human Erythrocytes Exposed to Phthalates and Their Metabolites Alter Antioxidant Enzyme Activity and Hemoglobin Oxidation. *Int J Mol Sci* 2020;21(12) doi: 10.3390/ijms21124480 [published Online First: 2020/07/01]

96. Teo AR, Choi H, Andrea SB, et al. Does Mode of Contact with Different Types of Social Relationships Predict Depression in Older Adults? Evidence from a Nationally Representative Survey. *Journal of the American Geriatrics Society* 2015;63(10):2014-22. doi: https://doi.org/10.1111/jgs.13667

97. Wang M-J, Yogeeswaran K, Andrews NP, et al. How Common Is Cyberbullying Among Adults? Exploring Gender, Ethnic, and Age Differences in the Prevalence of Cyberbullying. *Cyberpsychology, Behavior, and Social Networking* 2019;22(11):736-41. doi: 10.1089/cyber.2019.0146

Supplementary Table 1:
DISTRIBUTION ACROSS COUNTRIES, GENDERS AND AGE GROUPS.

COUNTRY	Total N	Age Group								Gender				
		18-24	25-34	35-44	45-54	55-64	65-74	75-84	85+	Female	Male	Non-binary/ Third Gender	Other/ Intersex	Prefer not to say
Algeria	7730	608	1061	1700	1683	1687	891	96	4	4382	3252	19	3	74
Argentina	13240	1149	837	1617	2400	3818	2728	618	73	8143	4890	89	0	118
Australia	10809	2286	877	719	1337	2340	2259	880	111	6396	4149	164	0	100
Belgium	2070	95	92	153	312	632	597	174	15	1272	781	8	0	9
Cameroon	2167	559	793	427	210	129	43	5	1	1229	922	5	0	11
Canada	13616	2313	1229	931	1483	2998	3175	1298	189	7962	5329	183	0	142
Chile	2094	127	58	154	308	782	548	114	3	1267	806	11	0	10
Colombia	10568	4475	1700	1351	1183	1221	545	88	5	6378	3951	116	0	123
Côte d'Ivoire	2057	472	668	495	272	114	31	4	1	922	1127	3	0	5
Ecuador	2119	693	196	223	327	450	198	29	3	1089	994	17	0	19
Egypt	4863	978	872	1438	913	489	155	16	2	2966	1824	4	3	66
France	3191	444	119	179	437	824	898	261	29	1930	1171	49	0	41
Guatemala	2481	557	346	469	485	442	163	18	1	1412	1040	10	0	19
India	36867	14186	6894	3333	4056	4715	3082	559	42	19224	17290	108	0	245
Iraq	2689	789	425	527	492	345	102	8	1	1638	1018	2	4	27
Ireland	3991	756	348	519	808	913	514	127	6	2247	1659	44	0	41
Mexico	12886	4618	1330	1449	1814	2311	1136	212	16	7500	5029	178	0	179
Morocco	3834	340	545	813	818	871	402	43	2	2194	1612	8	2	18
New Zealand	6397	1430	438	473	1077	1396	1106	434	43	3758	2475	113	0	51
Nigeria	6505	778	1076	1787	1601	1006	242	14	1	3479	2984	11	0	31
Peru	3372	936	207	272	465	825	544	115	8	1921	1392	23	0	36
Saudi Arabia	1643	654	256	282	253	144	47	5	2	1118	487	4	2	33
Singapore	3035	1006	402	338	508	461	270	48	2	1805	1133	46	0	51
South Africa	13945	1984	2045	2021	2412	2900	2014	530	39	8368	5428	83	0	66
Spain	8216	1087	423	926	1626	2577	1298	261	18	4788	3247	84	0	97
Tunisia	3777	265	344	638	815	965	617	124	9	2060	1689	6	0	22
United Kingdom	24706	4481	2597	2365	4696	5763	3625	1078	101	14441	9683	340	0	242
United States	37362	7553	3430	2667	3569	6289	8601	4525	728	21380	14693	813	0	476
Venezuela	13910	1729	1599	2325	3225	3263	1489	261	19	8056	5756	45	0	53
Yemen	2626	626	949	704	257	75	14	0	1	1642	958	1	2	23
Other Countries	7234													
TOTAL RECORDS								**270000**						

Supplementary Table 2:
ELEMENTS CAPTURED WITHIN EACH DETERMINANT CATEGORY

DETERMINANT	ELEMENTS CAPTURED
Demographics	Age; gender; country; language; educational attainment, employment status
Lifestyle	Frequency of sleeping well; frequency of exercise, frequency of in-person socializing with friends
Traumas and adversities	Experience of sexual abuse; cyberbullying; divorce; breakdown of romantic relationships; sudden or premature death of a family member; extreme poverty and homelessness; loss of a job; debilitating of life-threatening injury; loss due to natural disaster; participant or witness to war (see supplementary table 2 for full list)
Substances used	Tobacco; alcohol; cannabis ; vaping products; sedatives or sleeping pills; amphetamines; opioids
Medical Conditions	31 common medical conditions including: diabetes (type II); cancer, heart disease, hypertension, arthritis, migraine and traumatic brain injury

Supplementary Table 3:
ELEMENTS CAPTURED WITHIN EACH DETERMINANT CATEGORY

INDEX	CONTEXTUAL FACTOR	FEATURE
1	Basic Demographics	Age=18-24
2		Age=25-34
3		Age=35-44
4		Age=45-54
5		Age=55-64
6		Age=65-74
7		Age=75-84
8		Age=85+
9		Gender=Female
10		Gender=Male
11		Gender=Non-binary/Third Gender
12		Gender=Other/Intersex
13		Gender=Prefer not to say
14		Language
15		Country
16	Extended Demographics	Education=Bachelor's Degree
17		Education=High School
18		Education=Master's Degree
19		Education=Other
20		Education=Ph.D. or higher
21		Education=Prefer not to say
22		Education=Primary Education
23		Education=Professional Certificate
24		Education=Some High School
25		Education=Vocational certification
26		Employment=Employed /Self employed
27		Employment=Homemaker
28		Employment=Not able to work
29		Employment=Retired
30		Employment=Studying
31		Employment=Unemployed
32	Lifestyle Factors	In general, I get as much sleep as I need:=All of the time
33		In general, I get as much sleep as I need:=Hardly ever
34		In general, I get as much sleep as I need:=Most of the time
35		In general, I get as much sleep as I need:=Some of the time
36		In general, I get as much sleep as I need:=Sometimes
37		How regularly to you engage in physical exercise (30 minutes or more)?=Every day
38		How regularly to you engage in physical exercise (30 minutes or more)?=Few days a week
39		How regularly to you engage in physical exercise (30 minutes or more)?=Less than once a week
40		How regularly to you engage in physical exercise (30 minutes or more)?=Once a week
41		How regularly to you engage in physical exercise (30 minutes or more)?=Rarely/ Never
42		How regularly do you socialize with friends in person?=1-3 times a month
43		How regularly do you socialize with friends in person?=Once a week
44		How regularly do you socialize with friends in person?=Rarely/Never
45		How regularly do you socialize with friends in person?=Several days a week

46	Traumas and Adversities	Prolonged sexual abusel or severe sexual assault
47		Cyberbullying or online abuse
48		Loss of your job or livelihood leading to an inability to make ends meet
49		Divorce or family breakup
50		Sudden or premature death of a loved one
51		Caring for a child or partner with a major chronic disability or illness
52		I have not experienced any major trauma in my life
53		Breakdown of romantic relationship
54		Life threatening or debilitating injury or illness
55		Involvement or close witness to a war
56		Suffered a loss in a major fire, flood, earthquake, or natural disaster
57		Extreme poverty leading to homelessness and/or hunger
58		Prefer not to say (traumas)
59		Number of Traumas
60		My current job demands have become excessive or stressful
61		We have reduced household income
62		I have lost my job
63		My partner has lost their job
64		Other
65		I am unable to make ends meet for basic necessities
66		A close family member has died from Coronavirus
67		A close family member currently has or has had severe Coronavirus infection
68		I am isolated at home with my family
69		I currently have or have had mild Coronavirus infection
70		I am isolated at home on my own
71		Someone in my immediate family can't receive the critical healthcare support they need for an existing condition
72		I currently have or have had severe Coronavirus infection
73	Substance Use	Tobacco products
74		Alcoholic beverages
75		Cannabis
76		Vaping products
77		Sedatives or Sleeping Pills
78		Amphetamine type stimulants (e.g. speedl diet pillsl ecstasyl etc.)
79		Opioids
80		Melatonin
81	Medical Conditions	Type II Diabetes
82		Fibromyalgia
83		Liver disease/Cirrhosis
84		Hypertension
85		Irritable Bowel Syndrome
86		Heart disease
87		Inflammatory Bowel Disease / Crohn's disease
88		Arthritis
89		Psoriasis
90		Asthma
91		Migraines
92		Traumatic Brain Injury
93		Osteoporosis
94		Sleep apnea
95		Neuropathy
96		Cancer
97		Chronic Obstructive Pulmonary Disease (COPD)
98		HIV /AIDS
99		Kidney Disease
100		Polycystic ovaries
101		Narcolepsy
102		Chronic fatigue syndrome
103		Back problem
104		Epilepsy
105		Multiple sclerosis
106		Stroke
107		Herpes
108		Type 1 Diabetes
109	Occupation Features	High level of social interaction with other people
110		Knowledge/technical work
111		Teaching/training/mentoring others
112		Arts/Creative work
113		Administrative work
114		High level of physical activity
115		Caring for others
116		Requires extensive travel
117		Outdoors/Close to nature
118		Bus driver
119		Retail
120	Outcome variables	MHQ Score
121		MHQ Sign (positive or negative)

Supplementary Table 4:
NUMERICAL CODING OF ORDINAL SELECTIONS

DESCRIPTOR	CATEGORICAL VALUES	NUMERICAL VALUES
Education	Primary Education	1
	Some High School	2
	High School	3
	Bachelor's Degree	4
	Vocational certification	5
	Professional Certificate	5
	Master's Degree	6
	Ph.D. or higher	8
	Prefer not to say	null
	Other	null
Employment	Not able to work	1
	Unemployed	2
	Retired	3
	Homemaker	3
	Studying	4
	Employed /Self employed	5
Frequency of Good Sleep	Hardly ever	1
	Some of the time	2
	Sometimes	2
	Most of the time	3
	All the time	4
Frequency of Exercise	Rarely/Never	1
	Less than once a week	2
	Once a week	3
	Few days a week	4
	Every day	5
Frequency of In Person Socializing with friends	Rarely/Never	1
	1-3 times a month	2
	Once a week	3
	Several days a week	4

Supplementary Table 5:
PERFORMANCE METRICS OF CLASSIFICATION MODELS

Classification Method	AUC	CA	F1	Precision	Recall
Logistic Regression	0.83	0.80	0.79	0.77	0.78
Naïve Bayes	0.82	0.78	0.78	0.78	0.78
Gradient Boosting	0.84	0.80	0.79	0.79	0.80
Random Forest	0.80	0.78	0.77	0.77	0.78

AUC: area under the ROC curve; **CA:** classification accuracy

Supplementary Table 6:
LIFT CHARACTERISTICS OF GB CLASSIFICATION MODEL

	MHQ<0		MHQ≥0	
Top* % of Scores	Recall	Lift	Recall	Lift
5%	**0.85**	**2.7**	**0.99**	**1.43**
10%	**0.79**	**2.5**	**0.98**	**1.42**
15%	**0.74**	**2.4**	**0.97**	**1.40**
20%	**0.69**	**2.2**	**0.96**	**1.39**
50%	**0.67**	**2.1**	**0.90**	**1.30**

***** Lowest in the case of MHQ<0 and highest in the case of MHQ≥0;
MHQ: Mental Health Quotient

Supplementary Table 7:
PERFORMANCE METRICS OF REGRESSION MODELS

Regression method	MHQ<0		MHQ≥0		R-squared
	MHQ Points	% of Scale	MHQ Points	% of Scale	
Gradient Boosting	57	19%	46	15.3%	0.42
Random Forest	57	19%	46	15.3%	0.40
Linear Regression	55	18%	45	15.0%	0.44

RMSE: root mean squared error; **MAE:** mean absolute error

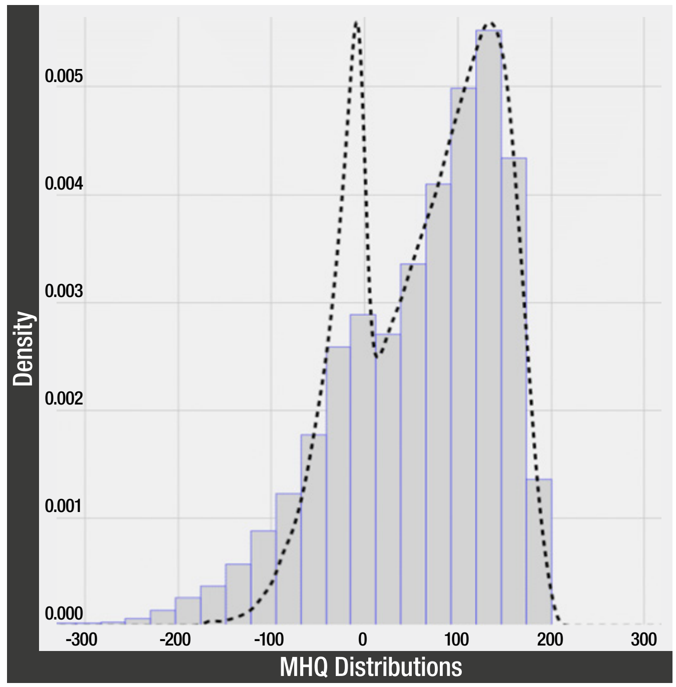
Supplementary Table 8:
CHARACTERISTICS BY CATEGORIES OF FEATURES USED

Determinant	AUC	CA	F1	Precision	Recall
Demographics1 (age; gender; country; language; education; employment)	0.775	0.767	0.744	0.746	0.767
Demographics1 + Lifestyle (exercise; sleep; socializing)	0.824	0.795	0.782	0.782	0.795
Demographics1 + Lifestyle + Traumas/adversities (type)	0.831	0.801	0.788	0.788	0.801
Demographics1 + Lifestyle + Traumas/adversities (type) + Medical conditions	0.833	0.802	0.790	0.790	0.802
Demographics1 + Lifestyle + Traumas/adversities (type) + Medical conditions + Substance use	0.837	0.804	0.792	0.792	0.804
Lifestyle (exercise; sleep; socializing)	0.727	0.753	0.715	0.723	0.753
Lifestyle + Traumas/ adversities (number)	0.736	0.757	0.715	0.728	0.757
Lifestyle + Traumas/ adversities (number) + Traumas/adversities (type)	0.754	0.765	0.734	0.742	0.765
Traumas/adversities (type)	0.642	0.745	0.679	0.709	0.745
Traumas/adversities (type) + Lifestyle (exercise)	0.685	0.750	0.686	0.722	0.750
Traumas/adversities (type) + Lifestyle (exercise; sleep)	0.733	0.759	0.716	0.732	0.759
Traumas/adversities (type) + Lifestyle (exercise; sleep; socializing)	0.754	0.766	0.734	0.742	0.766
Demographics2 (age; gender; country; language)	0.754	0.754	0.726	0.727	0.754
Demographics2 (age; gender; country; language) + Traumas/adversities (type)	0.774	0.769	0.744	0.747	0.769
Demographics2 (age; gender; country; language) + Education + Employment	0.775	0.767	0.745	0.746	0.767

AUC: area under the ROC curve; **CA:** classification accuracy

Supplementary Figure 1:
DISTRIBUTION OF MHQ SCORES.

Dotted line represents scores provided to users where negative scores have been rescaled to a smaller range.



Supplementary Figure 2:
SPEARMAN'S RANK CORRELATION BETWEEN CONTEXTUAL FACTORS

shows multiple inter-dependencies.
*** indicates significance of p<0•0001, X indicates p>0•05.**