Cross-sectional associations between physical activity and subdomains of mental health: A propensity score analysis in a global sample

Christopher Huong & Denver M. Y. Brown

**Introduction**

Suboptimal mental well-being and mental health problems have proven to be a significant burden on the global population. Approximately 29% of individuals have experienced a common mental disorder during their lifetime, with a 9.6%, 12.9%, and 10.7% lifetime prevalence for mood, anxiety, and substance-use disorders, respectively [1]. Global burden of mental disorders in 2019 is estimated at 125.3 million disability-adjusted life-years (DALY) – an increase from 80.8 million DALYs in 1990 [2]. Moreover, lost productivity due to poor mental health is estimated to cost the global economy $2.5T annually and is projected to more than double over the next decade, reaching $6T by 2030 [3]. The COVID-19 pandemic may also have conferred a toll on population mental health and wellbeing worldwide. Evidence from reviews is somewhat mixed, but findings suggest there has been an increase in depressive and anxiety symptoms most consistently in younger and female cohorts, [4-10].

Emerging evidence in addition to recent findings from the COVID-19 pandemic suggests certain age cohorts across the lifespan may be at greater risk for poor mental health and wellbeing than others [11]. For example, a recent report showed that younger adults living in the US consistently report the lowest scores on all domains of mental well-being, including happiness, health, meaning and purpose, character, social relationships, and financial stability, with a linear pattern of improvements in wellbeing observed with increased age [12]. These findings are in contrast to previous work that had demonstrated an inverted-U relationship between age and mental wellbeing in which mental wellbeing was lowest in middle adulthood [13]. Further, an estimated half of first mental disorder onsets occur by age 18 [14]. Evidently, more research is needed to better understand these trends from a global perspective, including a focus on protective factors that may moderate the relationship between age, mental health, and wellbeing.

Physical activity is one aspect of our lifestyle that has received increasing attention for its potential to promote mental wellbeing as well as assist in the prevention and management of mental health disorders. Observational and experimental evidence consistently shows a relationship between higher levels of physical activity and benefits across various mental health outcomes, including, but not limited to depression [15-20], anxiety [21-23], and general mental health [24-26]. For example, using non-parametric matching in a large cross-sectional dataset, Chekroud, et al. (27) showed that engaging in physical exercise in the past month was associated with 43.2% lower self-reported days with poor mental health over that period. Despite these robust findings, particularly for depression and anxiety, physical activity remains an underutilized treatment tool among clinical practitioners [28]. The inclusion of physical activity – albeit as an alternative and/or complementary treatment – in several recent national guidelines for the treatment of depression holds promise for promoting greater uptake in primary care services [29-32].

While evidence supports the importance of physical activity for preventing and/or reducing a range of mental health problems and improving overall wellbeing, for some other mental symptoms and disorders (e.g., bipolar disorder, schizophrenia), the evidence is less suggestive of benefits, or remains unclear [33-38]. More recently, however, researchers have observed considerable overlap in symptomology across the most commonly classified mental health disorders, which illustrates the complexity and heterogeneous nature of mental health as a construct [39, 40]. Such knowledge has sparked the development of novel measures – the Mental Health Quotient (MHQ) for example – to address these considerations. Yet, due to their recency, these instruments have seldom been utilized.

One promising avenue in which comprehensive measures of mental health can be applied is in studies examining the specific aspects of mental health that physical activity may especially favor. At present, there is a paucity of literature in this area – to our knowledge only one study has investigated symptom-level effects [41]. Findings stemming from such studies have the potential to improve precision in the promotion of mental wellbeing and treatment of mental health problems and therefore may have important implications for clinical outcomes [42-44]. Determining whether physical activity is associated with differential effects for certain aspects of mental health across the lifespan is but one fruitful area of inquiry to pursue for the purpose of improving population-level mental health and well-being.

The purpose of the present study was to estimate the treatment effect of regular physical activity engagement on overall mental health in a large global sample, as well as various subdomains of mental health, while statistically accounting for a range of observed covariates using a machine learning technique underutilized in exercise psychology and behavioral medicine.

**Study Sample and Data Collection**

This cross-sectional study used data from the Mental Health Million (MHM) project, an on-going study with the purpose of assessing global mental wellbeing through administration of the Mental Health Quotient. The sample for our present study included 341,956 participants from 229 countries who completed the MHQ from December 31st, 2021 to Oct 14th, 2022; This start of this period coincided with the launch of Version 3 of the MHQ. Additional information concerning the MHM project and recruitment strategy may be found elsewhere [45]. This study involved secondary analysis of existing data and therefore Institutional Research Ethics Board approval was not required.

**Measures:**

**Mental Health:**

The MHQ is a 47-item instrument designed to assess a comprehensive range of common attributes found across widely used existing mental health assessment tools in a single questionnaire to estimate overall mental wellbeing and functioning in the population. Items were developed by consolidation of 170 symptoms coded from 126 commonly used assessment tools covering depression, anxiety, bipolar disorder, ADHD, post-traumatic stress disorder, obsessive-compulsive disorder, addiction, schizophrenia, eating disorders and autism spectrum disorder. The MHQ is unique from other psychiatric tools in that the items assessed the level of functioning and impact on one’s life associated with each mental health element, as opposed to frequency, duration, or severity of symptoms. The questionnaire took an average of 14 minutes for participants to complete.

Responses from the 47 items were used to compute the MHQ, which represents an overall mental health and wellbeing score. The MHQ originally ranged from -100 to +200, however, the lower limit was recently expanded to -166 to improve the distribution of scores that previously demonstrated a floor effect at the lower bound. Scores on the MHQ can be classified into six levels of functioning, with negative scores indicating clinical risk and positive scores representing normal range: Clinical (≤-50), At Risk (-50 to <0), Enduring (0 to <50), Managing (50 to <100), Succeeding (100 to <150) and Thriving (150 to 200). To compute the overall score, individual item responses were weighted to reflect the nonlinearity of risk associated with increases in symptom severity, as well as the differential risk associated with different symptoms (e.g., suicidal thoughts vs irritability).

In addition to the overall MHQ score, scores for six broad subcategories of mental health and wellbeing were computed: Core Cognition (ability for executive functioning), Adaptability and Resilience (decision making, creativity, and tolerance to change), Mood and Outlook (ability to effectively regulate ones emotions), Drive and Motivation (ability to achieve goals in the face of obstacles), Social Self (social functioning), and Mind-Body (physical functioning and psychosomatic health). Subcategory scores ranged from -100 to +200, and were computed by a weighted average of scores from 10 to 24 relevant symptom items based on a review of cognitive and brain functioning models [45].

The MHQ has demonstrated high sample reliability when four randomly selected and demographically similar samples were compared on response distributions (p = 0.99), and resulting MHQ distribution (p = 0.18) [46]. Internal consistency was demonstrated with conceptually similar items having higher correlations than unsimilar items. A subset of participants which took the MHQ twice at least 3 days apart showed a test-retest reliability of r = 0.84. Validity was assessed by asking a subset of participants additional questions concerning days missed from work and normal activities in the past month. Those who were employed and scored an overall MHQ between 175 to 200 missed on average 0.2 days of work in the past month, while those employed who scored between -75 to -100 missed an average of 9.3 days of work.

**Physical Activity**: Participants responded to single item that asked: “How regularly do you engage in physical exercise (30 minutes or more)?” Response options included “Rarely/never”; “Less than once a week”; “Once a week”; “Few days a week”; and “Every day”. In line with Chekroud, et al. (27), responses to the physical activity frequency item were recoded into binary groups with participants who reported “Rarely/never” coded as the control (inactive; no exposure to physical activity), and all other responses coded the treated (physically active; exposure to physical activity).

**Covariates**: To adjust for potential confounders, the following covariates were included in our statistical models: age, biological sex, gender identity, ethnicity, educational attainment, employment status, relationship status, frequency of adequate sleep, frequency of socializing, diagnosis of a medical condition (Y/N), whether they are currently seeking mental health treatment (Y/N), and whether they reported a significant traumatic childhood or adult experience (Y/N). Data inspection revealed considerable missingness for ethnicity (84.2%) and gender identity (98.5%), and therefore these variables were dropped from our analyses. All items in which participants responded “Prefer not to say” were recoded as missing.

**Statistical Analysis:**

All data preprocessing and statistical analyses were done using the statistical software R version 4.1.2 [47] and RStudio (Version 2022.07.2). First, distributions of covariates were balanced between the non-exposure and exposure groups using propensity score weights estimated with generalized boosted modeling (GBM) [48, 49], and implemented in the R Package *WeightIt* [50]. The propensity score is defined as “the conditional probability of assignment to a particular treatment given a vector of observed covariates” [51]. Weighting was preferred over matching procedures for the purpose of preserving sample size. Propensity scores weights were computed based on the Average Treatment effect on the Treated (ATT) estimand, which is used to estimate the hypothetical effect (i.e., counterfactual) of physical activity exposure on the control (i.e. inactive) group by making the control similar to the treated across covariate distribution. Treatment effects estimated from propensity score adjustment are unbiased when the strong ignorability assumption is met (i.e., when there are no unobserved confounders, and all observed confounders are included in the model) [51].

GBM is a non-parametric iterative machine learning method which, as implemented in the present study, combines boosting (the sequential combination of weak learners to improve predictions by adapting the errors of the previous model) and regression trees (the weak learners) to generate a smoothed function of estimated propensity scores. This method automatically accommodates non-linearity and complex interactions, and has been shown in previous studies to outperform traditional parametric models such as logistic regression [48, 52, 53].

Several tuning parameters were selected to achieve covariate balancing, as suggested by McCaffrey, et al. (48). The Bernoulli distribution was chosen for the loss function the boosted model was to minimize, as our treatment variable was coded as a binary exposure. The number of trees was determined by minimizing the average standardized absolute mean difference in the covariates. The maximum number of trees was set to 10,000 by default, and increased to 20,000 if covariate balancing was not achieved. Missing data was handled by surrogate splitting as described in the *WeightIt* R package documentation [50]. Though unlikely to significantly improve the performance of our procedure [54] weights above 99% were winsorized to reduce potential bias from extreme values. Diagnostics were used to ensure covariates were adequately balanced by assessing the weighted absolute standardized difference in means of covariates between treatment and control group, the weighted Kolmogorov-Smirnov (KS) statistic.

For our main analysis, we estimated the ATT for physical activity on seven outcomes: overall MHQ score, and its six subdomains: Core Cognition, Adaptability and Resilience, Mood and Outlook, Drive and Motivation, Social Self, and Mind-Body. Using the same procedures, we also tested whether age moderates the relationship between physical activity and the seven mental health outcomes. In all models, participants were nested within country to account for potential clustering effects. ATTs are presented as beta coefficients with 95% confidence intervals.

In case the inferences of the main analyses were biased due to model misspecification or handling of missing data [55, 56], we performed several sensitivity analyses. First, we computed propensity score weighted linear regression models that included further adjustment for the full covariate set to allow for doubly robust estimation [57]. Second, we estimated ATTs using multiple imputation (MI) and covariate balancing propensity score (CBPS) weighting, which may outperform GBM if there is a non-complex relationship between treatment and outcome [58]. This procedure is described in greater detail in the Supplementary Materials (Section A). Doubly robust estimation was also computed for these models by including the full covariate set in the multiply imputed, CBPS weighted linear regression models. Lastly, we computed a (non-covariate balanced) linear regression model to examine the mean difference in MHQ scores between the physical activity exposure and non-exposure groups, adjusted for covariates. Akin to our main analyses, participants were nested within country for all models.

**Results:**

**Descriptive statistics**

After dropping two cases, the final sample included 341,956 participants, and was predominantly female (55.3%), educated (47.5% with a bachelor’s or graduate degree), employed (47.8%), married (42.5%), and physically active (60.4%). The sample was also representative across the adult life span (18-24 and 55-64 were the most common age ranges at 18.91% and 18.50% of the sample respectively). The mean score for the MHQ was 67.93 ± 72.70 SD. Full descriptive statistics for the sample demographics characteristics, physical activity, and mental health can be found in Table 1.

After adjusting for propensity score weighting, the effective sample for the untreated (inactive) group was reduced to 89,597.54 (66.11% of unadjusted), yielding an overall effective sample size of 296,028.5 (86.57% of original sample). Diagnostics indicated that covariate balance was successfully achieved after GBM weighting procedures were implemented (Supplementary Materials, Section B).

**Physical activity**

Propensity score weighted models demonstrated physical activity was associated with significantly (p < 0.01) higher MHQ scores (B = 18.45; 95% CI: 15.52-21.37), which coincided with a small effect size (Cohen’s *d* = 0.26). Physical activity was also associated with significantly higher scores for each of the six MHQ subdomains: Core Cognition (~~B =~~ 16.44; 95% CI: 13.77-19.10; *d* = 0.25), Adaptability and Resilience (B =18.14; 95% CI: 15.40-20.88; *d* = 0.27), Drive and Motivation (B = 15.75; 95% CI: 12.62-18.87; *d* = 0.24), Mood and Outlook (B = 16.32; 95% CI: 13.44-19.20; *d* = 0.24), Social Self (B = 14.03; 95% CI: 11.12-16.93; *d* = 0.19), and Mind-Body (B = 19.81; 95% CI: 17.23-22.36; *d* = 0.32).

**Physical Activity by Age Interaction**

The propensity score weighted moderation models revealed a significant (p < 0.05) physical activity by age interaction for overall MHQ scores (B = -2.04; 95% CI: -3.18 - -0.90), Core Cognition (B = -2.09; 95% CI: -3.10 - -1.08), Mood and Outlook (B = -2.04; 95% CI: -3.20 - -0.89), Drive and Motivation (B = -2.15; 95% CI: -3.42 - -0.88), and Mind-Body (B = -2.20; 95% CI: -3.49 - -0.91), but not Adaptability and Resilience (B = -1.16; 95% CI: -2.30 - -0.02, p = 0.0457) and Social Self (B = -1.05; 95% CI: -2.20 – 0.09, = 0.0708). These results indicate younger age cohorts experience more favorable benefits for mental health and most of its domains from physical activity. Significant main effects for overall MHQ scores were observed for physical activity ( B = 25.64; 95% CI: 20.42 – 30.86) and age (B = 14.94; 95% CI: 13.43 – 16.44), as well as for the subdomains (shown in Table 3).

**Sensitivity analyses**

Our sensitivity analyses showed convergence of the estimated treatment effects of physical activity on mental health across each of the alternative statistical analysis techniques employed when compared to the main GBM results (B = 18.45; 95% CI: 15.52 – 21.37): doubly robust GBM (B = 18.07; 95% CI: 15.95 – 20.19), MI + CBPS (B = 18.04), doubly robust CBPS (B = 17.87; 95% CI: 17.43 – 18.32), non-covariate balanced linear regression model (B = 18.07; 95% CI: 15.95 – 20.19).

**Discussion**

The present study demonstrates a significant association between self-reported physical activity and Mental Health Quotient score, a novel measure of overall mental health which broadly captures clinical symptoms as well as aspects of wellbeing, summarizing them on a single dimension ranging from clinical risk to thriving. The effect was robust after adjusting for potential selection bias to mimic random-assignment of the treatment (i.e., physical activity) by assigning weights to participants based on the probability they were physical active or inactive, given their full set of observed covariates.

MHQ validation + why improving it matters

* Favorable benefits of physical activity on overall mental health as well as several subdomains (Main effect of PA)
  + How does this compare to chekroud
  + Supports general body of literature demonstrating the benefits of PA
  + Ours results compared to RCTs (compare effect sizes?)
* PE strongest effect on Adaptability and Resilience, lowest on Social Self (is this in line with current literature? Comments on different exercise modalities?) (How PA differed across each subdomain – where are the strongest/weakest effects – why may this be)
  + Implications for precision medicine – PA stands to provide greatest benefits for some outcomes, maybe other behaviors provide greater benefits for others – to be determined in future use of the instrument (MHQ)
* Relatively poor MH in younger ages, confirming Chen 2022 (main effect of age)
* Interaction effects exist, stronger benefits for PA among younger adults.
  + This may be due to them having more to gain due to poorer mental health
  + Comment on implications for population level interventions potentially
* Propensity weighting methods allow us to make stronger causal inferences compared to common regression models (using more advanced techniques)
* Limitations
* Conclusion

There are several limitations with the current study. Firstly, unlike randomization, propensity score weighting does not adjust for unobserved covariates (Joffe & Rosenbaum, 1999). An unbiased treatment effect assumes that all potential confounders are observed, which is unlikely to be the case in any observational study. Adjusting for partially observed confounders may reduce bias, but it is unknown to what extent.

Limitations:

Self-reported physical activity behavior – can introduce bias and recall errors

Cross-sectional design – cannot infer causality

Convenience sampling – not truly globally representative despite over 200 countries included

Strong ignorability assumption; propensity scoring assumes all potentially confounding covariates are observed

Dropped ethnicity and gender identity

Self reported PA

Cross sectional

Volunteer online survey

Ordinal nature of PA questions = difficult to assess exact non-linear effects

1. Steel Z, Marnane C, Iranpour C, et al.: The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *International journal of epidemiology.* 2014, *43:*476-493.

2. Collaborators GMD: Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Psychiatry.* 2022, *9:*137-150.

3. Health TLG: Mental health matters. *The Lancet. Global Health.* 2020, *8:*e1352.

4. Samji H, Wu J, Ladak A, et al.: Mental health impacts of the COVID‐19 pandemic on children and youth–a systematic review. *Child and adolescent mental health.* 2022, *27:*173-189.

5. Patel K, Robertson E, Kwong AS, et al.: Psychological distress before and during the COVID-19 pandemic among adults in the United Kingdom based on coordinated analyses of 11 longitudinal studies. *JAMA Network open.* 2022, *5:*e227629-e227629.

6. Panchal U, Salazar de Pablo G, Franco M, et al.: The impact of COVID-19 lockdown on child and adolescent mental health: systematic review. *European child & adolescent psychiatry.* 2021*:*1-27.

7. Prati G, Mancini AD: The psychological impact of COVID-19 pandemic lockdowns: a review and meta-analysis of longitudinal studies and natural experiments. *Psychological medicine.* 2021, *51:*201-211.

8. Robinson E, Sutin AR, Daly M, Jones A: A systematic review and meta-analysis of longitudinal cohort studies comparing mental health before versus during the COVID-19 pandemic in 2020. *Journal of Affective Disorders.* 2022, *296:*567-576.

9. Kauhanen L, Wan Mohd Yunus WMA, Lempinen L, et al.: A systematic review of the mental health changes of children and young people before and during the COVID-19 pandemic. *European child & adolescent psychiatry.* 2022*:*1-19.

10. Santomauro DF, Herrera AMM, Shadid J, et al.: Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet.* 2021, *398:*1700-1712.

11. Oswalt SB, Lederer AM, Chestnut-Steich K, et al.: Trends in college students’ mental health diagnoses and utilization of services, 2009–2015. *Journal of American college health.* 2020, *68:*41-51.

12. Chen Y, Cowden RG, Fulks J, Plake JF, VanderWeele TJ: National data on age gradients in well-being among US adults. *JAMA psychiatry.* 2022, *79:*1046-1047.

13. Blanchflower DG, Oswald AJ: Is well-being U-shaped over the life cycle? *Social science & medicine.* 2008, *66:*1733-1749.

14. Solmi M, Radua J, Olivola M, et al.: Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Molecular psychiatry.* 2022, *27:*281-295.

15. Gordon BR, McDowell CP, Hallgren M, et al.: Association of efficacy of resistance exercise training with depressive symptoms: meta-analysis and meta-regression analysis of randomized clinical trials. *JAMA psychiatry.* 2018, *75:*566-576.

16. Kvam S, Kleppe CL, Nordhus IH, Hovland A: Exercise as a treatment for depression: a meta-analysis. *Journal of Affective Disorders.* 2016, *202:*67-86.

17. 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.

18. Schuch F, Vancampfort D, Firth J, et al.: Physical activity and sedentary behavior in people with major depressive disorder: a systematic review and meta-analysis. *Journal of Affective Disorders.* 2017, *210:*139-150.

19. Schuch FB, Vancampfort D, Richards J, et al.: Exercise as a treatment for depression: a meta-analysis adjusting for publication bias. *Journal of psychiatric research.* 2016, *77:*42-51.

20. Schuch FB, Vancampfort D, Rosenbaum S, et al.: Exercise improves physical and psychological quality of life in people with depression: A meta-analysis including the evaluation of control group response. *Psychiatry research.* 2016, *241:*47-54.

21. Aylett E, Small N, Bower P: Exercise in the treatment of clinical anxiety in general practice–a systematic review and meta-analysis. *BMC health services research.* 2018, *18:*1-18.

22. Biddle SJ, Ciaccioni S, Thomas G, Vergeer I: Physical activity and mental health in children and adolescents: An updated review of reviews and an analysis of causality. *Psychology of Sport and Exercise.* 2019, *42:*146-155.

23. McDowell CP, Dishman RK, Gordon BR, Herring MP: Physical activity and anxiety: a systematic review and meta-analysis of prospective cohort studies. *American journal of preventive medicine.* 2019, *57:*545-556.

24. Firth J, Solmi M, Wootton RE, et al.: A meta‐review of “lifestyle psychiatry”: the role of exercise, smoking, diet and sleep in the prevention and treatment of mental disorders. *World Psychiatry.* 2020, *19:*360-380.

25. Marquez DX, Aguiñaga S, Vásquez PM, et al.: A systematic review of physical activity and quality of life and well-being. *Translational behavioral medicine.* 2020, *10:*1098-1109.

26. Rodriguez-Ayllon M, Cadenas-Sánchez C, Estévez-López F, et al.: Role of physical activity and sedentary behavior in the mental health of preschoolers, children and adolescents: a systematic review and meta-analysis. *Sports Medicine.* 2019, *49:*1383-1410.

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. *The Lancet Psychiatry.* 2018, *5:*739-746.

28. Ekkekakis P: Why Is Exercise Underutilized in Clinical Practice Despite Evidence It Is Effective? Lessons in Pragmatism From the Inclusion of Exercise in Guidelines for the Treatment of Depression in the British National Health Service. *Kinesiology Review.* 2020, *10:*29-50.

29. *Depression in adults: treatment and management*. London: National Institute for Health and Care Excellence (NICE), 2022.

30. Ravindran AV, Balneaves LG, Faulkner G, et al.: Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: section 5. Complementary and alternative medicine treatments. *The Canadian Journal of Psychiatry.* 2016, *61:*576-587.

31. Stubbs B, Vancampfort D, Hallgren M, et al.: EPA guidance on physical activity as a treatment for severe mental illness: a meta-review of the evidence and Position Statement from the European Psychiatric Association (EPA), supported by the International Organization of Physical Therapists in Mental Health (IOPTMH). *European Psychiatry.* 2018, *54:*124-144.

32. Malhi GS, Bassett D, Boyce P, et al.: Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. *Australian & New Zealand Journal of Psychiatry.* 2015, *49:*1087-1206.

33. Firth J, Cotter J, Elliott R, French P, Yung AR: A systematic review and meta-analysis of exercise interventions in schizophrenia patients. *Psychological medicine.* 2015, *45:*1343-1361.

34. Dauwan M, Begemann MJ, Heringa SM, Sommer IE: Exercise improves clinical symptoms, quality of life, global functioning, and depression in schizophrenia: a systematic review and meta-analysis. *Schizophrenia bulletin.* 2016, *42:*588-599.

35. Brokmeier LL, Firth J, Vancampfort D, et al.: Does physical activity reduce the risk of psychosis? A systematic review and meta-analysis of prospective studies. *Psychiatry research.* 2020, *284:*112675.

36. Brondino N, Rocchetti M, Fusar‐Poli L, et al.: A systematic review of cognitive effects of exercise in depression. *Acta Psychiatrica Scandinavica.* 2017, *135:*285-295.

37. Ashdown-Franks G, Firth J, Carney R, et al.: Exercise as medicine for mental and substance use disorders: a meta-review of the benefits for neuropsychiatric and cognitive outcomes. *Sports Medicine.* 2020, *50:*151-170.

38. Melo MCA, Daher EDF, Albuquerque SGC, de Bruin VMS: Exercise in bipolar patients: a systematic review. *Journal of Affective Disorders.* 2016, *198:*32-38.

39. Borsboom D, Cramer AO, Schmittmann VD, Epskamp S, Waldorp LJ: The small world of psychopathology. *PloS one.* 2011, *6:*e27407.

40. Newson JJ, Pastukh V, Thiagarajan TC: Poor separation of clinical symptom profiles by DSM-5 disorder criteria. *Frontiers in psychiatry.* 2021, *12*.

41. Murri MB, Ekkekakis P, Menchetti M, et al.: Physical exercise for late-life depression: effects on symptom dimensions and time course. *Journal of Affective Disorders.* 2018, *230:*65-70.

42. Uher R, Perlis R, Henigsberg N, et al.: Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. *Psychological medicine.* 2012, *42:*967-980.

43. Iniesta R, Malki K, Maier W, et al.: Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of psychiatric research.* 2016, *78:*94-102.

44. Fried EI, Nesse RM: Depression sum-scores don’t add up: why analyzing specific depression symptoms is essential. *BMC medicine.* 2015, *13:*1-11.

45. Newson JJ, Thiagarajan TC: Assessment of population well-being with the Mental Health Quotient (MHQ): development and usability study. *JMIR Mental Health.* 2020, *7:*e17935.

46. Newson JJ, Pastukh V, Thiagarajan TC: Assessment of Population Well-being With the Mental Health Quotient: Validation Study. *JMIR Mental Health.* 2022, *9:*e34105.

47. RCoreTeam: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., 2022.

48. McCaffrey DF, Ridgeway G, Morral AR: Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods.* 2004, *9:*403.

49. Friedman JH: Greedy function approximation: a gradient boosting machine. *Annals of statistics.* 2001*:*1189-1232.

50. Greifer N: WeightIt: weighting for covariate balance in observational studies. *R package version 0.10.* 2020, *2*.

51. Rosenbaum PR, Rubin DB: The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983, *70:*41-55.

52. Tu C: Comparison of various machine learning algorithms for estimating generalized propensity score. *Journal of Statistical Computation and Simulation.* 2019, *89:*708-719.

53. Lee BK, Lessler J, Stuart EA: Improving propensity score weighting using machine learning. *Statistics in medicine.* 2010, *29:*337-346.

54. Lee BK, Lessler J, Stuart EA: Weight trimming and propensity score weighting. *PloS one.* 2011, *6:*e18174.

55. Cham H, West SG: Propensity score analysis with missing data. *Psychological methods.* 2016, *21:*427.

56. Coffman DL, Zhou J, Cai X: Comparison of methods for handling covariate missingness in propensity score estimation with a binary exposure. *BMC medical research methodology.* 2020, *20:*1-14.

57. Funk MJ, Westreich D, Wiesen C, et al.: Doubly robust estimation of causal effects. *American journal of epidemiology.* 2011, *173:*761-767.

58. Setodji CM, McCaffrey DF, Burgette LF, Almirall D, Griffin BA: The right tool for the job: Choosing between covariate balancing and generalized boosted model propensity scores. *Epidemiology (Cambridge, Mass.).* 2017, *28:*802.

**Table 1. Population descriptives**

|  | Total | 18-24 | 25-34 | 35-44 | 45-54 | 55-64 | 65-74 | 75-84 | 85+ |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ( = 341956) | (N= 64648) | (N= 47249) | (N= 55241) | (N= 57259) | (N= 63113) | (N= 40881) | (N= 12029) | (N= 1536) |
| Sex |  |  |  |  |  |  |  |  |  |
| Female | 189226 (55.3%) | 39831 (61.6%) | 27344 (57.9%) | 31203 (56.5%) | 30890 (53.9%) | 32729 (51.9%) | 20256 (49.5%) | 6141 (51.1%) | 832 (54.2%) |
| Male | 148624 (43.5%) | 23206 (35.9%) | 19427 (41.1%) | 23584 (42.7%) | 25815 (45.1%) | 29827 (47.3%) | 20299 (49.7%) | 5807 (48.3%) | 659 (42.9%) |
| Other/Intersex | 819 (0.2%) | 268 (0.4%) | 63 (0.1%) | 92 (0.2%) | 163 (0.3%) | 136 (0.2%) | 63 (0.2%) | 14 (0.1%) | 20 (1.3%) |
| Missing | 3287 (1.0%) | 1343 (2.1%) | 415 (0.9%) | 362 (0.7%) | 391 (0.7%) | 421 (0.7%) | 263 (0.6%) | 67 (0.6%) | 25 (1.6%) |
| Education |  |  |  |  |  |  |  |  |  |
| Associate’s Degree | 21259 (6.2%) | 3921 (6.1%) | 2512 (5.3%) | 2643 (4.8%) | 3174 (5.5%) | 4381 (6.9%) | 3398 (8.3%) | 1093 (9.1%) | 137 (8.9%) |
| Bachelor’s Degree | 105724 (30.9%) | 17800 (27.5%) | 19818 (41.9%) | 20404 (36.9%) | 18197 (31.8%) | 17066 (27.0%) | 9753 (23.9%) | 2455 (20.4%) | 231 (15.0%) |
| Graduate Degree | 56764 (16.6%) | 3386 (5.2%) | 8580 (18.2%) | 11690 (21.2%) | 11725 (20.5%) | 11671 (18.5%) | 7423 (18.2%) | 2038 (16.9%) | 251 (16.3%) |
| High School | 77434 (22.6%) | 23451 (36.3%) | 7871 (16.7%) | 9293 (16.8%) | 10880 (19.0%) | 12985 (20.6%) | 9321 (22.8%) | 3155 (26.2%) | 478 (31.1%) |
| Less than High School | 34040 (10.0%) | 6836 (10.6%) | 2902 (6.1%) | 4720 (8.5%) | 5674 (9.9%) | 7249 (11.5%) | 4811 (11.8%) | 1620 (13.5%) | 228 (14.8%) |
| Other | 16246 (4.8%) | 3578 (5.5%) | 2248 (4.8%) | 2552 (4.6%) | 2733 (4.8%) | 2997 (4.7%) | 1649 (4.0%) | 412 (3.4%) | 77 (5.0%) |
| Vocational Certification | 21823 (6.4%) | 1922 (3.0%) | 2613 (5.5%) | 3054 (5.5%) | 3842 (6.7%) | 5487 (8.7%) | 3762 (9.2%) | 1043 (8.7%) | 100 (6.5%) |
| Missing | 8666 (2.5%) | 3754 (5.8%) | 705 (1.5%) | 885 (1.6%) | 1034 (1.8%) | 1277 (2.0%) | 764 (1.9%) | 213 (1.8%) | 34 (2.2%) |
| Employment |  |  |  |  |  |  |  |  |  |
| Employed /Self employed | 163401 (47.8%) | 12630 (19.5%) | 29040 (61.5%) | 39198 (71.0%) | 41026 (71.6%) | 33075 (52.4%) | 7391 (18.1%) | 960 (8.0%) | 81 (5.3%) |
| Homemaker | 31570 (9.2%) | 2260 (3.5%) | 5383 (11.4%) | 8198 (14.8%) | 7162 (12.5%) | 5874 (9.3%) | 2124 (5.2%) | 478 (4.0%) | 91 (5.9%) |
| Not able to work | 7210 (2.1%) | 1212 (1.9%) | 806 (1.7%) | 966 (1.7%) | 1444 (2.5%) | 2193 (3.5%) | 473 (1.2%) | 81 (0.7%) | 35 (2.3%) |
| Retired | 61333 (17.9%) | 102 (0.2%) | 93 (0.2%) | 317 (0.6%) | 1924 (3.4%) | 17194 (27.2%) | 30038 (73.5%) | 10380 (86.3%) | 1285 (83.7%) |
| Studying | 48583 (14.2%) | 40925 (63.3%) | 4306 (9.1%) | 1476 (2.7%) | 1101 (1.9%) | 574 (0.9%) | 149 (0.4%) | 30 (0.2%) | 22 (1.4%) |
| Unemployed | 29859 (8.7%) | 7519 (11.6%) | 7621 (16.1%) | 5086 (9.2%) | 4602 (8.0%) | 4203 (6.7%) | 706 (1.7%) | 100 (0.8%) | 22 (1.4%) |
| Relationship Status |  |  |  |  |  |  |  |  |  |
| Divorced/ Separated | 32484 (9.5%) | 876 (1.4%) | 2137 (4.5%) | 4738 (8.6%) | 6849 (12.0%) | 9677 (15.3%) | 6535 (16.0%) | 1536 (12.8%) | 136 (8.9%) |
| In a cohabiting relationship | 15388 (4.5%) | 1318 (2.0%) | 2861 (6.1%) | 3170 (5.7%) | 3062 (5.3%) | 3022 (4.8%) | 1576 (3.9%) | 345 (2.9%) | 34 (2.2%) |
| In a relationship | 32066 (9.4%) | 12729 (19.7%) | 7347 (15.5%) | 3650 (6.6%) | 3146 (5.5%) | 3103 (4.9%) | 1669 (4.1%) | 390 (3.2%) | 32 (2.1%) |
| Married/Civil Partnership | 145166 (42.5%) | 2022 (3.1%) | 14235 (30.1%) | 31501 (57.0%) | 33622 (58.7%) | 35037 (55.5%) | 22295 (54.5%) | 5897 (49.0%) | 557 (36.3%) |
| Other | 2313 (0.7%) | 803 (1.2%) | 470 (1.0%) | 439 (0.8%) | 339 (0.6%) | 208 (0.3%) | 46 (0.1%) | 5 (0.0%) | 3 (0.2%) |
| Single (never married or in a civil partnership) | 85117 (24.9%) | 42004 (65.0%) | 18279 (38.7%) | 9084 (16.4%) | 6653 (11.6%) | 5955 (9.4%) | 2583 (6.3%) | 508 (4.2%) | 51 (3.3%) |
| Widowed | 14744 (4.3%) | 139 (0.2%) | 158 (0.3%) | 557 (1.0%) | 1490 (2.6%) | 3860 (6.1%) | 4895 (12.0%) | 2984 (24.8%) | 661 (43.0%) |
| Missing | 14678 (4.3%) | 4757 (7.4%) | 1762 (3.7%) | 2102 (3.8%) | 2098 (3.7%) | 2251 (3.6%) | 1282 (3.1%) | 364 (3.0%) | 62 (4.0%) |
| Socialize Frequency |  |  |  |  |  |  |  |  |  |
| Rarely/Never | 86212 (25.2%) | 14926 (23.1%) | 14063 (29.8%) | 16429 (29.7%) | 16052 (28.0%) | 15390 (24.4%) | 7446 (18.2%) | 1683 (14.0%) | 223 (14.5%) |
| 1-3 times a month | 79457 (23.2%) | 13115 (20.3%) | 11989 (25.4%) | 12973 (23.5%) | 13545 (23.7%) | 15310 (24.3%) | 9591 (23.5%) | 2655 (22.1%) | 279 (18.2%) |
| Once a week | 63389 (18.5%) | 11242 (17.4%) | 8055 (17.0%) | 9235 (16.7%) | 9847 (17.2%) | 12275 (19.4%) | 9352 (22.9%) | 3019 (25.1%) | 364 (23.7%) |
| Several days a week | 112898 (33.0%) | 25365 (39.2%) | 13142 (27.8%) | 16604 (30.1%) | 17815 (31.1%) | 20138 (31.9%) | 14492 (35.4%) | 4672 (38.8%) | 670 (43.6%) |
| Adequate Sleep Frequency |  |  |  |  |  |  |  |  |  |
| Hardly ever | 42262 (12.4%) | 10071 (15.6%) | 6175 (13.1%) | 7270 (13.2%) | 7113 (12.4%) | 7180 (11.4%) | 3498 (8.6%) | 836 (6.9%) | 119 (7.7%) |
| Some of the time | 117966 (34.5%) | 25492 (39.4%) | 18234 (38.6%) | 21289 (38.5%) | 20129 (35.2%) | 19465 (30.8%) | 10499 (25.7%) | 2569 (21.4%) | 289 (18.8%) |
| Most of the time | 140096 (41.0%) | 22792 (35.3%) | 17866 (37.8%) | 21383 (38.7%) | 23592 (41.2%) | 27803 (44.1%) | 19799 (48.4%) | 6106 (50.8%) | 755 (49.2%) |
| All of the time | 41632 (12.2%) | 6293 (9.7%) | 4974 (10.5%) | 5299 (9.6%) | 6425 (11.2%) | 8665 (13.7%) | 7085 (17.3%) | 2518 (20.9%) | 373 (24.3%) |
| Medical Diagnosis |  |  |  |  |  |  |  |  |  |
| No | 280712 (82.1%) | 54877 (84.9%) | 40813 (86.4%) | 47336 (85.7%) | 47419 (82.8%) | 49690 (78.7%) | 31090 (76.1%) | 8488 (70.6%) | 999 (65.0%) |
| Yes | 53164 (15.5%) | 7584 (11.7%) | 5438 (11.5%) | 6702 (12.1%) | 8628 (15.1%) | 12133 (19.2%) | 8917 (21.8%) | 3286 (27.3%) | 476 (31.0%) |
| Missing | 8080 (2.4%) | 2187 (3.4%) | 998 (2.1%) | 1203 (2.2%) | 1212 (2.1%) | 1290 (2.0%) | 874 (2.1%) | 255 (2.1%) | 61 (4.0%) |
| Mental Health Treatment in Past Year |  |  |  |  |  |  |  |  |  |
| No | 287518 (84.1%) | 51770 (80.1%) | 38632 (81.8%) | 46434 (84.1%) | 48454 (84.6%) | 53656 (85.0%) | 36239 (88.6%) | 10916 (90.7%) | 1417 (92.3%) |
| Yes | 50606 (14.8%) | 12028 (18.6%) | 8183 (17.3%) | 8221 (14.9%) | 8181 (14.3%) | 8732 (13.8%) | 4192 (10.3%) | 979 (8.1%) | 90 (5.9%) |
| Missing | 3832 (1.1%) | 850 (1.3%) | 434 (0.9%) | 586 (1.1%) | 624 (1.1%) | 725 (1.1%) | 450 (1.1%) | 134 (1.1%) | 29 (1.9%) |
| Experienced Childhood Trauma |  |  |  |  |  |  |  |  |  |
| No | 131004 (38.3%) | 16518 (25.6%) | 15289 (32.4%) | 21660 (39.2%) | 24223 (42.3%) | 27401 (43.4%) | 19235 (47.1%) | 5958 (49.5%) | 720 (46.9%) |
| Yes | 210952 (61.7%) | 48130 (74.4%) | 31960 (67.6%) | 33581 (60.8%) | 33036 (57.7%) | 35712 (56.6%) | 21646 (52.9%) | 6071 (50.5%) | 816 (53.1%) |
| Experienced Adult Trauma |  |  |  |  |  |  |  |  |  |
| No | 89087 (26.1%) | 17408 (26.9%) | 12573 (26.6%) | 16663 (30.2%) | 15786 (27.6%) | 15432 (24.5%) | 9027 (22.1%) | 1999 (16.6%) | 199 (13.0%) |
| Yes | 252869 (73.9%) | 47240 (73.1%) | 34676 (73.4%) | 38578 (69.8%) | 41473 (72.4%) | 47681 (75.5%) | 31854 (77.9%) | 10030 (83.4%) | 1337 (87.0%) |
| Physical Activity |  |  |  |  |  |  |  |  |  |
| Inactive | 135525 (39.6%) | 27949 (43.2%) | 23082 (48.9%) | 27178 (49.2%) | 24080 (42.1%) | 20054 (31.8%) | 9828 (24.0%) | 2883 (24.0%) | 471 (30.7%) |
| Active | 206431 (60.4%) | 36699 (56.8%) | 24167 (51.1%) | 28063 (50.8%) | 33179 (57.9%) | 43059 (68.2%) | 31053 (76.0%) | 9146 (76.0%) | 1065 (69.3%) |
| MHQ |  |  |  |  |  |  |  |  |  |
| Mean (SD) | 67.9 (72.7) | 21.2 (69.6) | 45.4 (69.2) | 64.6 (68.4) | 80.2 (67.6) | 92.6 (65.5) | 103 (59.7) | 111 (55.1) | 111 (63.7) |
| Median [Min, Max] | 79.1 [-166, 200] | 2.10 [-166, 200] | 45.2 [-166, 200] | 72.6 [-166, 200] | 93.5 [-166, 200] | 108 [-166, 200] | 117 [-166, 200] | 124 [-166, 200] | 125 [-166, 200] |
| Core Cognition |  |  |  |  |  |  |  |  |  |
| Mean (SD) | 81.4 (67.0) | 39.5 (63.2) | 61.0 (64.7) | 78.6 (64.7) | 92.3 (63.0) | 104 (60.4) | 113 (54.7) | 119 (51.3) | 116 (58.3) |
| Median [Min, Max] | 93.5 [-100, 200] | 25.4 [-100, 200] | 63.3 [-100, 200] | 88.8 [-100, 200] | 106 [-100, 200] | 118 [-100, 200] | 125 [-100, 200] | 130 [-100, 200] | 130 [-100, 200] |
| Adaptability and Resilience |  |  |  |  |  |  |  |  |  |
| Mean (SD) | 89.8 (67.4) | 50.7 (67.5) | 70.6 (66.7) | 88.3 (64.6) | 101 (62.7) | 110 (61.3) | 118 (55.2) | 125 (49.7) | 123 (55.4) |
| Median [Min, Max] | 105 [-100, 200] | 53.7 [-100, 200] | 81.7 [-100, 200] | 102 [-100, 200] | 116 [-100, 200] | 124 [-100, 200] | 132 [-100, 200] | 137 [-100, 200] | 134 [-100, 200] |
| Drive and Motivation |  |  |  |  |  |  |  |  |  |
| Mean (SD) | 83.4 (66.6) | 43.5 (64.7) | 62.6 (65.2) | 80.4 (63.9) | 94.3 (62.4) | 105 (60.1) | 114 (53.8) | 118 (49.7) | 114 (57.1) |
| Median [Min, Max] | 94.3 [-100, 200] | 36.8 [-100, 200] | 66.7 [-100, 200] | 89.7 [-100, 200] | 106 [-100, 200] | 120 [-100, 200] | 126 [-100, 200] | 129 [-100, 200] | 124 [-100, 200] |
| Mood and Outlook |  |  |  |  |  |  |  |  |  |
| Mean (SD) | 67.2 (71.1) | 24.4 (64.8) | 44.1 (66.3) | 62.0 (67.1) | 78.1 (67.6) | 91.2 (66.2) | 102 (61.5) | 111 (57.2) | 114 (61.7) |
| Median [Min, Max] | 73.7 [-100, 200] | -1.30 [-100, 200] | 35.6 [-100, 200] | 65.4 [-100, 200] | 88.8 [-100, 200] | 105 [-100, 200] | 116 [-100, 200] | 124 [-100, 200] | 129 [-100, 200] |
| Social Self |  |  |  |  |  |  |  |  |  |
| Mean (SD) | 70.7 (76.6) | 23.1 (69.5) | 50.2 (72.5) | 69.7 (72.8) | 83.5 (72.8) | 94.3 (71.8) | 103 (68.2) | 112 (64.6) | 116 (67.3) |
| Median [Min, Max] | 83.9 [-100, 200] | -3.40 [-100, 200] | 46.8 [-100, 200] | 80.6 [-100, 200] | 101 [-100, 200] | 115 [-100, 200] | 123 [-100, 200] | 131 [-100, 200] | 134 [-100, 200] |
| Mind-Body |  |  |  |  |  |  |  |  |  |
| Mean (SD) | 73.4 (64.7) | 44.6 (64.3) | 57.7 (63.8) | 68.7 (63.8) | 80.5 (62.9) | 90.1 (60.8) | 98.5 (55.5) | 102 (52.0) | 99.3 (57.0) |
| Median [Min, Max] | 84.1 [-100, 200] | 45.5 [-100, 200] | 64.5 [-100, 200] | 78.6 [-100, 200] | 92.7 [-100, 200] | 102 [-100, 200] | 109 [-100, 200] | 111 [-100, 200] | 108 [-100, 200] |

**Table 2. Treatment effect of physical activity from GBM-estimated propensity scores**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Outcome | M | 95% CI | SE | Cohen’s d | p-value |
| Overall Mental Health Quotient | 18.45 | 15.52 – 21.37 | 1.48 | 0.26 | < 0.001 |
| Core Cognition | 16.44 | 13.77 – 19.10 | 1.35 | 0.25 | < 0.001 |
| Adaptability and Resilience | 18.14 | 15.40 – 20.88 | 1.39 | 0.27 | < 0.001 |
| Drive and Motivation | 15.75 | 12.62 – 18.87 | 1.59 | 0.24 | < 0.001 |
| Mood and Outlook | 16.32 | 13.44 – 19.20 | 1.46 | 0.24 | < 0.001 |
| Social Self | 14.03 | 11.12 – 16.93 | 1.47 | 0.19 | < 0.001 |
| Mind-Body | 19.81 | 17.25 – 22.36 | 1.30 | 0.32 | < 0.001 |

**Table 3. Physical activity X Age Interaction**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | MHQ | Core Cognition | Adaptability and Resilience | Mood and Outlook | Drive and Motivation | Social Self | Mind-Body |
| Physical Activity | 25.64 | 24.00 | 21.63 | 23.33 | 23.81 | 16.89 | 28.32 |
| SE | 2.65 | 2.32 | 2.66 | 2.59 | 2.87 | 2.74 | 2.75 |
| p-value | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| Age | 14.94 | 13.73 | 11.83 | 14.59 | 13.07 | 13.97 | 9.70 |
| SE | 0.77 | 0.64 | 0.78 | 0.68 | 0.90 | 0.85 | 0.80 |
| p-value | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 | < 0.001 |
| PA X Age | -2.04 | -2.09 | -1.16 | -2.05 | -2.15 | -1.05 | -2.20 |
| SE | 0.58 | 0.51 | 0.58 | 0.59 | 0.64 | 0.58 | 0.66 |
| p-value | < 0.001 | < 0.001 | 0.0457 | < 0.001 | 0.001 | 0.071 | < 0.001 |

**Table 4. Sensitivity analysis. Reported treatment effects and standard errors**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | MHQ | Core Cognition | Adaptability and Resilience | Mood and Outlook | Drive and Motivation | Social Self | Mind-Body |
| GBM | 18.45 (1.48) |  |  |  |  |  |  |
| MI + GBM |  |  |  |  |  |  |  |
| MI + CBPS | 18.04 (0.27) |  |  |  |  |  |  |
| Multivariable Regression | 18.07 (1.07) |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |