Cross-sectional associations between regular physical exercise engagement and domains of mental wellbeing in a large global sample

Christopher Huong & Denver M. Y. Brown

(another title option?)

Estimating the treatment effect of regular exercise on mental wellbeing with propensity score weighting using generalized boosted models

Using a machine learning method to estimate the treatment effect of regular exercise on mental wellbeing

**Introduction**

Mental health conditions 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 (Steel et al., 2014). 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 (Collaborators, 2022). 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 (Health, 2020).

Recent epidemiological reviews of global samples indicates stagnation in the overall prevalence of mental disorders (Richter et al., 2019), despite an increase in the prevalence of depression (Moreno-Agostino et al., 2021). Together, these studies suggest that current strategies which seek to prevent and/or reduce mental problems are ineffective (Jorm et al., 2017). The COVID-19 pandemic may also have conferred a toll on population mental health and wellbeing worldwide. Evidence from reviews is somewhat mixed, suggesting there has been an increase in depressive and anxiety symptoms most consistently in younger cohorts, with other studies finding evidence for psychological resilience (Panchal et al., 2021; Patel et al., 2022; Prati & Mancini, 2021; Robinson et al., 2022; Samji et al., 2022).

Emerging evidence suggests certain age cohorts across the adult lifespan may be at greater risk for poor mental health and wellbeing than others (Oswalt et al., 2020). For example, a recent report showed that younger adults living in the US consistently report the lowest scores on all domains of 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 (Chen et al., 2022). 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 (Blanchflower & Oswald, 2008). 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 and mental health and wellbeing.

An increasing body of literature continues to establish effects of physical activity on mental health as well. Observational and experimental evidence consistently shows a relationship between higher levels of physical activity and benefits across various mental health outcomes (for depression see Gordon et al., 2018; Kvam et al., 2016; Pearce et al., 2022; Schuch et al., 2017; Schuch, Vancampfort, Richards, et al., 2016; Schuch, Vancampfort, Rosenbaum, et al., 2016; for anxiety see Aylett et al., 2018; Biddle et al., 2019; McDowell et al., 2019; for general mental health see Firth et al., 2020; Marquez et al., 2020; Rodriguez-Ayllon et al., 2019). For example, using non-parametric matching in a large cross-sectional dataset, Chekroud et al. (2018) showed that reporting having exercised in the past month was associated with 43.2% lower self-reported days with poor mental health. Despite these robust findings, particularly for depression and anxiety, physical activity remains an underutilized treatment tool among clinical practitioners (Ekkekakis, 2020). While this evidence further supports the importance of physical activity for preventing and/or reducing a range of mental health problems and improving overall wellbeing, certain mental health disorders and symptoms have received limited attention to date. For some other mental disorders and symptoms, the evidence is less suggestive of benefits, or remains unclear (Ashdown-Franks et al., 2020; Brokmeier et al., 2020; Brondino et al., 2017; Dauwan et al., 2016; Firth et al., 2015; Melo et al., 2016).

Recent work has also suggested that mental health is a complex and heterogenous construct in which there is considerable overlap in symptomology across the most commonly classified disorders (Borsboom et al., 2011; Newson et al., 2021). Novel measures such as the Mental Health Quotient (MHQ) have been developed to address these considerations, but due to their recency, have received limited attention. There is also a paucity of literature on the specific aspects of mental health that physical activity may especially favor, which only one previous study to our knowledge investigating symptom-level effects (Murri et al., 2018). Improved precision in the treatment target may have important implications for clinical outcomes (Fried & Nesse, 2015; Iniesta et al., 2016; Uher et al., 2012). Physical activity may be associated with differential effects across the lifespan, particularly as it relates to certain aspects of mental wellbeing.

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

**Study Sample and Data Collection**

Our study used a cross-sectional dataset 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 XXX to XXX. Additional information concerning the MHM project and recruitment strategy may be found elsewhere (Newson & Thiagarajan, 2020). This study involved secondary analysis of existing data and therefore Institutional Research Ethics Board approval was not required.

**Measures:**

**Mental Wellbeing:**

The MHQ is a 47-item voluntary online survey 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 were computed into an overall mental wellbeing score, originally ranging from -100 to +200. Recently, the lower limit was expanded to -166 to accommodate a floor effect. Scores are binned 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 wellbeing score, scores for six broad subcategories of mental 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 (Newson & Thiagarajan, 2020).

The MHQ 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). 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 (Newson et al., 2022).

**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”.

**Covariates**: To adjust for potential confounders, we selected as covariates age, biological sex, gender identity, ethnicity, educational attainment, employment status, relationship status, frequency of adequate sleep, frequency of socializing, diagnosis of 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). All items in which participants responded “Prefer not to say” were recoded as missing. Participants were further nested by country in the analysis to account for potential clustering effects.

**Statistical Analysis:**

All data preprocessing and statistical analyses were done using the statistical software R version 4.1.2 (RCoreTeam, 2022). As the items used to calculate the overall MHQ score, as well as exercise frequency, were required by the questionnaire, only some covariate items which were not required showed significant missingness, with ethnicity and gender identity showing the highest (84.2% and 98.5% respectively). These were thus we dropped it from further analysis. Answer responses which included “Prefer not to say” were recoded to missing. In line with Chekroud et al. (2018), responses to physical exercise (PE) frequency were recoded into binary groups, with “Rarely/Never” indicating the control (no exposure to exercise), and all other responses indicating the treated (exposure to exercise).

Distributions of covariates were balanced between the non-exposure and exposure groups using propensity score weights estimated with generalized boosted modeling (GBM) (McCaffrey et al., 2004), and implemented in the R-Package WeightIt (Greifer, 2020). The propensity score is defined as “the conditional probability of assignment to a particular treatment given a vector of observed covariates” (Rosenbaum & Rubin, 1983) with weighting procedures preferred over matching to preserve the size of the sample. Propensity scores were converted into weights based on the Average Treatment effect on the Treated estimand, which is used to estimate the hypothetical effect of exercise exposure on the control (i.e. non-exercise) group, and is unbiased when the strong ignorability assumption is met (Rosenbaum & Rubin, 1983). The GBM is a non-parametric iterative machine learning method which utilizes regression trees to generate predicted values (i.e. of propensity scores). As suggested by McCaffrey et al. (2004) the number of iterations was determined by minimizing the average standardized absolute mean difference in the covariates. The number of trees was set to 10,000 by default, and increased to 20,000 if covariate balancing was not achieved. The GBM handles missing data by the surrogate splitting method as described in the WeightIt R-package documentation (Greifer, 2020). Though unlikely to significantly improve performance of inferences for our estimation procedure (Lee et al., 2011) weights above 99% were trimmed (i.e. set to 99%) to reduce potential bias from extreme values. Diagnostics were used to ensure covariates were adequately balanced. Our main analysis estimated treatment effects of exercise on seven outcomes: overall MHQ score, and the six broad subcategories of wellbeing. Using the same procedures, we also computed a model testing for moderation of age on PE and overall MHQ score as well as the subcategories.

In case the inferences of the main analysis are biased due to model misspecification or handling of missing data, we performed several sensitivity analyses. First, we conducted a doubly robust estimator of the GBM estimations by combining the propensity weights with the outcome linear regression (Funk et al., 2011). Second, we estimated treatment effects 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 (Setodji et al., 2017) (more on this procedure in Supplementary A). We also computed a double robust estimator using MI + CBPS weighting. Lastly, we computed mean difference of MHQ between no exercise and exercise groups, adjusting for covariates and nesting by country.

**Results:**

Descriptive statistics of the sample can be found in Supplementary B. Diagnostics indicated that control and treatment groups were not significantly different across all levels of covariates after weighting, indicating successful balancing. Mean levels of MHQ in the full sample was 67.93 (+/- 72.70). In the main model, self-reported PE was associated with an 18.45 (95% CI: 15.52-21.37) increase in MHQ (Cohen’s d = 0.26), 16.44 (13.77-19.10) increase in Core Cognition (d = 0.25), 18.14 (15.40-20.88) increase in Adaptability and Resilience (d = 0.27), 15.75 (12.62-18.87) increase in Drive and Motivation (d = 0.24), 16.32 (13.44-19.20) increase in Mood and Outlook (d = 0.24), 14.03 (11.12-16.93) increase in Social Self (d = 0.19), and a 19.81 (17.23-22.36) increase in Mind-Body (d = 0.32) subcategories.

The propensity score weighted moderation model revealed significant age\*PE interactions on MHQ (B = -2.04), Core Cognition (B = -2.09), Adaptability and Resilience (B = -1.16), Mood and Outlook (B = -2.04), Drive and Motivation (B = -2.15), and Mind-Body (B = -2.20), indicating younger age cohorts benefit more from physical exercise across overall wellbeing and all subcategories, excluding Social Self. Significant main effects of age are also observed across all subcategories (B = 25.64, 13.43-16.44 for MHQ).

The sensitivity analyses shows convergence of estimated treatment effects, suggesting robustness of the estimation to different model specifications (main GBM = 18.45, double robust GBM = 18.07, MI + CBPS = 18.04, double robust CBPS = 17.87, adjusted mean difference = 18.07).

**Discussion**

• Propensity weighting methods allow us to make stronger causal inferences compared to common regression models.

• Ours results compared to RCTs (compare effect sizes?)

• Relatively poor MH in younger ages, confirming Chen 2022

• PE strongest effect on Adaptability and Resilience, lowest on Social Self (is this in line with current literature? Comments on different exercise modalities?)

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.

Limitations:

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

Ashdown-Franks, G., Firth, J., Carney, R., Carvalho, A. F., Hallgren, M., Koyanagi, A., Rosenbaum, S., Schuch, F. B., Smith, L., & Solmi, M. (2020). Exercise as medicine for mental and substance use disorders: a meta-review of the benefits for neuropsychiatric and cognitive outcomes. *Sports Medicine*, *50*(1), 151-170.

Blanchflower, D. G., & Oswald, A. J. (2008). Is well-being U-shaped over the life cycle? *Social science & medicine*, *66*(8), 1733-1749.

Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS one*, *6*(11), e27407.

Brokmeier, L. L., Firth, J., Vancampfort, D., Smith, L., Deenik, J., Rosenbaum, S., Stubbs, B., & Schuch, F. B. (2020). Does physical activity reduce the risk of psychosis? A systematic review and meta-analysis of prospective studies. *Psychiatry research*, *284*, 112675.

Brondino, N., Rocchetti, M., Fusar‐Poli, L., Codrons, E., Correale, L., Vandoni, M., Barbui, C., & Politi, P. (2017). A systematic review of cognitive effects of exercise in depression. *Acta Psychiatrica Scandinavica*, *135*(4), 285-295.

Chekroud, S. R., Gueorguieva, R., Zheutlin, A. B., Paulus, M., Krumholz, H. M., Krystal, J. H., & Chekroud, A. M. (2018). 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*, *5*(9), 739-746.

Chen, Y., Cowden, R. G., Fulks, J., Plake, J. F., & VanderWeele, T. J. (2022). National data on age gradients in well-being among US adults. *JAMA psychiatry*, *79*(10), 1046-1047.

Collaborators, G. M. D. (2022). 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*, *9*(2), 137-150.

Dauwan, M., Begemann, M. J., Heringa, S. M., & Sommer, I. E. (2016). Exercise improves clinical symptoms, quality of life, global functioning, and depression in schizophrenia: a systematic review and meta-analysis. *Schizophrenia bulletin*, *42*(3), 588-599.

Ekkekakis, P. (2020). 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*, *10*(1), 29-50.

Firth, J., Cotter, J., Elliott, R., French, P., & Yung, A. R. (2015). A systematic review and meta-analysis of exercise interventions in schizophrenia patients. *Psychological medicine*, *45*(7), 1343-1361.

Fried, E. I., & Nesse, R. M. (2015). Depression sum-scores don’t add up: why analyzing specific depression symptoms is essential. *BMC medicine*, *13*(1), 1-11.

Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly robust estimation of causal effects. *American journal of epidemiology*, *173*(7), 761-767.

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

Health, T. L. G. (2020). Mental health matters. *The Lancet. Global Health*, *8*(11), e1352.

Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N., Dernovsek, M. Z., Souery, D., & Stahl, D. (2016). Combining clinical variables to optimize prediction of antidepressant treatment outcomes. *Journal of psychiatric research*, *78*, 94-102.

Jorm, A. F., Patten, S. B., Brugha, T. S., & Mojtabai, R. (2017). Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry*, *16*(1), 90-99.

Lee, B. K., Lessler, J., & Stuart, E. A. (2011). Weight trimming and propensity score weighting. *PloS one*, *6*(3), e18174.

McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological methods*, *9*(4), 403.

Melo, M. C. A., Daher, E. D. F., Albuquerque, S. G. C., & de Bruin, V. M. S. (2016). Exercise in bipolar patients: a systematic review. *Journal of Affective Disorders*, *198*, 32-38.

Moreno-Agostino, D., Wu, Y.-T., Daskalopoulou, C., Hasan, M. T., Huisman, M., & Prina, M. (2021). Global trends in the prevalence and incidence of depression: a systematic review and meta-analysis. *Journal of Affective Disorders*, *281*, 235-243.

Murri, M. B., Ekkekakis, P., Menchetti, M., Neviani, F., Trevisani, F., Tedeschi, S., Latessa, P. M., Nerozzi, E., Ermini, G., & Zocchi, D. (2018). Physical exercise for late-life depression: effects on symptom dimensions and time course. *Journal of Affective Disorders*, *230*, 65-70.

Newson, J. J., Pastukh, V., & Thiagarajan, T. C. (2021). Poor separation of clinical symptom profiles by DSM-5 disorder criteria. *Frontiers in psychiatry*, *12*.

Newson, J. J., Pastukh, V., & Thiagarajan, T. C. (2022). Assessment of Population Well-being With the Mental Health Quotient: Validation Study. *JMIR Mental Health*, *9*(4), e34105.

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

Oswalt, S. B., Lederer, A. M., Chestnut-Steich, K., Day, C., Halbritter, A., & Ortiz, D. (2020). Trends in college students’ mental health diagnoses and utilization of services, 2009–2015. *Journal of American college health*, *68*(1), 41-51.

Panchal, U., Salazar de Pablo, G., Franco, M., Moreno, C., Parellada, M., Arango, C., & Fusar-Poli, P. (2021). The impact of COVID-19 lockdown on child and adolescent mental health: systematic review. *European child & adolescent psychiatry*, 1-27.

Patel, K., Robertson, E., Kwong, A. S., Griffith, G. J., Willan, K., Green, M. J., Di Gessa, G., Huggins, C. F., McElroy, E., & Thompson, E. J. (2022). 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*, *5*(4), e227629-e227629.

Prati, G., & Mancini, A. D. (2021). The psychological impact of COVID-19 pandemic lockdowns: a review and meta-analysis of longitudinal studies and natural experiments. *Psychological medicine*, *51*(2), 201-211.

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

Richter, D., Wall, A., Bruen, A., & Whittington, R. (2019). Is the global prevalence rate of adult mental illness increasing? Systematic review and meta‐analysis. *Acta Psychiatrica Scandinavica*, *140*(5), 393-407.

Robinson, E., Sutin, A. R., Daly, M., & Jones, A. (2022). 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*, *296*, 567-576.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41-55.

Samji, H., Wu, J., Ladak, A., Vossen, C., Stewart, E., Dove, N., Long, D., & Snell, G. (2022). Mental health impacts of the COVID‐19 pandemic on children and youth–a systematic review. *Child and adolescent mental health*, *27*(2), 173-189.

Setodji, C. M., McCaffrey, D. F., Burgette, L. F., Almirall, D., & Griffin, B. A. (2017). The right tool for the job: Choosing between covariate balancing and generalized boosted model propensity scores. *Epidemiology (Cambridge, Mass.)*, *28*(6), 802.

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: a systematic review and meta-analysis 1980–2013. *International journal of epidemiology*, *43*(2), 476-493.

Uher, R., Perlis, R., Henigsberg, N., Zobel, A., Rietschel, M., Mors, O., Hauser, J., Dernovsek, M., Souery, D., & Bajs, M. (2012). Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. *Psychological medicine*, *42*(5), 967-980.