

# **Associations between physical activity and subcategories of mental health: A propensity score analysis among a global sample of 341,956 adults**

## **Introduction**

Physical activity is a lifestyle factor that has received increasing attention for its potential to promote mental well-being as well as assist in the prevention and management of mental health disorders. Observational and experimental evidence has consistently shown a relationship between higher levels of physical activity and benefits across various mental health outcomes, including, but not limited to depression (Gordon et al., 2018; Pearce et al., 2022; Schuch, Vancampfort, Richards, et al., 2016), anxiety (Aylett et al., 2018; McDowell et al., 2019), and general mental health and well-being (Chekroud et al., 2018). Despite these findings, particularly for depression and anxiety, physical activity remains an underutilized treatment option among clinical practitioners (Ekkekakis, 2020). 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 (Kendrick et al., 2022; Malhi et al., 2015; Stubbs et al., 2018).

Emerging evidence also suggests younger age cohorts of adults may be at the greatest risk for poor mental health and well-being (Keyes et al., 2019; Oswald et al., 2020). For example, a recent study 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 well-being 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 well-being in which mental well-being was lowest in middle adulthood (Blanchflower & Oswald,

2008). Further, the onset of an estimated half of first mental disorders occurs by age 18, which speaks to the pervasiveness of mental health challenges experienced when transitioning into early adulthood (Solmi et al., 2022).

More recently, researchers have demonstrated that there is considerable overlap in symptomatology across the most commonly classified mental health disorders, which illustrates the complexity and heterogeneity of mental health constructs (Borsboom et al., 2011; Newson et al., 2021), and the need for symptom-specific considerations in clinical interventions (Boschloo et al., 2019; Fried, 2017; Iniesta et al., 2016). Such considerations have sparked the development of novel measures – such as the Mental Health Quotient (MHQ) – 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 determining whether physical activity is associated with benefits to specific aspects of mental health, especially in different age groups. To our knowledge, only one study has investigated the symptom-specific effects of physical activity (Murri et al., 2018), finding that compared to antidepressants only, antidepressants combined with exercise showed higher improvements in affective but not somatic symptoms in older depressed patients.

Although randomized controlled trials are the gold-standard for assessing the efficacy of treatments, certain interventions may be burdensome for researchers to implement (e.g., personalized exercise programs), resulting in smaller sample sizes, which may be especially problematic when targeting smaller subpopulations (e.g., older adults with depression). Thus, utilizing observational studies to establish preliminary associations is a useful way to explore more precise intervention strategies. An increasingly common method to adjust for selection bias in observational data is to balance covariates across treatment groups using propensity score

weighting, where the propensity score is defined as “the conditional probability of assignment to a particular treatment given a vector of observed covariates” (Rosenbaum & Rubin, 1983), and used as weights in a regression model. This quasi-experimental method has demonstrated utility in various fields such as preventative medicine (Pollack et al., 2010), dentistry (Burgette et al., 2016), community psychology (Lanza et al., 2013), and medicine (Fessler et al., 2019; Shin et al., 2021). For example, in comparing the effects of different types of exercise (e.g., cycling, sports, running) on mental health burden, (Chekroud et al., 2018) used propensity weighting to balance the distribution of covariates (e.g., age, race, gender) across exercise types.

The purpose of the present study was to estimate the effect of physical activity engagement on overall mental health, as well as various subcategories of mental health, and explore this relationship across age groups in a large global cross-sectional sample.

## Methods

The reporting of the methods and results in this study adhere to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines for cross-sectional studies. (Von Elm et al., 2007)

**Study Sample and Data Collection.** This cross-sectional study used data from the Global Mind Project (GMP; formally the Mental Health Million Project), an on-going online study with the purpose of assessing global mental well-being through administration of the Mental Health Quotient. The sample for our present study included 341,956 participants from 229 countries and territories who completed the GMP survey between December 31<sup>st</sup>, 2021 and October 14<sup>th</sup>, 2022. The start of this period coincided with the launch of Version 3 of the MHQ. Additional information concerning the GMP and recruitment strategy may be found elsewhere

(Newson & Thiagarajan, 2020). This study involved secondary analysis of existing non-identifiable 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 well-being 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 items assess level of functioning and impact on one's life associated with each mental health element, and 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 score of mental health and well-being. The MHQ originally ranged from -100 to +200, however, the lower limit was recently expanded to -166 to accommodate a floor effect. 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 ( $\leq -50$ ), 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 well-being were computed: Core Cognition (e.g., executive functioning and other cognitive

processes), Adaptability and Resilience (e.g., creativity and flexibility), Mood and Outlook (e.g., emotional regulation, optimism), Drive and Motivation (e.g., sustained interest, persistence), Social Self (e.g., maintaining relationships, self-image), and Mind-Body Connection (e.g., physical functioning, psychosomatic wellbeing). 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. Further details on the development, full descriptions, and psychometric properties of the MHQ can be found elsewhere (Newson et al., 2022; Newson & Thiagarajan, 2020),

**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., 2018), 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 as the treated (physically active; exposure to physical activity).

**Covariates.** To adjust for potential confounders, the following covariates were considered for inclusion in our analysis: 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%) due to only having been included on surveys for individuals who reported residing in certain countries, and therefore these variables were

excluded. All responses in which participants answered “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 (R Core Team, 2022) and RStudio (Version 2022.07.2). First, distributions of covariates were balanced between the non-exposure (i.e., Inactive) and exposure (i.e., Physically Active) groups using propensity score weights estimated with generalized boosted modeling (GBM) (Friedman, 2001; McCaffrey et al., 2004), and implemented in the R Package *WeightIt* (Greifer, 2022). 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 Control (ATC) estimand, which is used to estimate the hypothetical average treatment effect on those who did not receive the treatment. In other words, it is the expected effect of physical activity on those in the sample who are inactive, which would help inform the practical question of whether mental health practitioners should encourage physical activity in their sedentary patients. This research question is relevant for mental health practitioners given that individuals with mental health disorders have been shown to be more sedentary and less active than population norms (Vancampfort et al., 2017). Treatment effects estimated with 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) (Rosenbaum & Rubin, 1983).

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 (Lee et al., 2010; McCaffrey et al., 2004; Tu, 2019). Several tuning parameters were selected to achieve covariate balancing, as suggested by McCaffrey et al (McCaffrey et al., 2004). The Bernoulli distribution was chosen for the loss function the boosted model was to minimize, as our treatment variable was coded as binary. 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 (Greifer, 2022). Though unlikely to significantly improve the performance of our procedure (Lee et al., 2011) 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.

Propensity weights were fed into a regression model to estimate the ATC for physical activity on seven outcomes: overall MHQ score, and its six subcategories, Core Cognition, Adaptability and Resilience, Mood and Outlook, Drive and Motivation, Social Self, and Mind-Body Connection. To explore whether associations of physical activity and mental health differs by age, we performed the same analysis on each age group (18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-84, 85+ years of age) for the MHQ and its six subcategories. In all models, participants were nested within country to account for potential clustering effects.

Several sensitivity analyses were performed to determine whether the inferences of the main analyses were biased due to model misspecification or handling of missing data (Cham &

West, 2016; Coffman et al., 2020). First, we computed propensity score weighted regression models that included further adjustment for the full covariate set to allow for doubly robust estimation (Funk et al., 2011). Second, instead of handling covariate missingness by surrogate splitting, we first implemented multiple imputation (MI) before the GBM estimation of propensity scores. Third, we estimated ATCs using 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). MI and CBPS are 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 or GBM-weighted regression models. Akin to our main analyses, participants were nested within country for all models.

## Results

**Descriptive statistics.** After dropping two cases due to Arabic responses not translating properly, the final sample included 341,956 participants from 227 countries and territories, and was predominantly female (55.3%; 43.5% male, 0.2% other/intersex), post-secondary educated (47.5% with a bachelor's or graduate degree; 32.6% high school or less), employed (47.8%; 32.1% studying or retired, 8.7% unemployed), married (42.5%; 24.9% single), and physically active to some degree (60.4%). The sample was also representative across the adult lifespan (18-24 and 55-64 were the most common age ranges selected at 18.91% and 18.50% of the sample, respectively). The mean score for the MHQ was  $67.93 \pm 72.70$  SD, which would be considered "Managing" as per the MHQ's six levels of functioning. Full descriptive statistics for the sample demographic characteristics, covariates, physical activity, MHQ, and the six subcategories for each age group can be found in Supplementary Materials Table 1.



After adjusting for propensity score weighting, the effective sample for the treated (active) group was reduced to 140,633.8 (68.13% of unadjusted), yielding an overall effective sample size of 276,158.8 (80.76% of original sample). The effective sample size is the “approximately the number of observations from a simple random sample that yields an estimate with sampling variation equal to the sampling variation obtained with the weighted comparison observation”, and can be interpreted as a conservative lower bound for the adjusted size of the weighted sample (Ridgeway et al., 2022). Diagnostics indicated that covariate balance was successfully achieved after GBM and CBPS weighting procedures were implemented (see Supplementary Materials, Section B).

**Physical Activity.** Propensity score weighted models demonstrated physical activity was associated with greater MHQ scores (ATC = 17.86; 95% CI: 15.07-20.64), which coincided with a small effect size calculated using standard deviations from the unweighted data (standardized mean difference (SMD) = 0.25). Physical activity was also associated greater scores for each of the six MHQ subcategories: Core Cognition (ATC = 16.33; 95% CI: 13.87-18.78; SMD = 0.25), Adaptability and Resilience (ATC = 17.57; 95% CI: 14.83-20.31; SMD = 0.26), Drive and Motivation (ATC = 15.86; 95% CI: 12.87-18.86; SMD = 0.24), Mood and Outlook (ATC = 15.27; 95% CI: 12.53-18.01; SMD = 0.22), Social Self (ATC = 13.02; 95% CI: 10.18-15.85; SMD = 0.17), and Mind-Body Connection (ATC = 19.25; 95% CI: 16.66-21.84; SMD = 0.31).

**Physical Activity and Age.** Figure 1 shows the estimated effects of PA on overall MHQ scores and each subcategory by each age group. Briefly, overall trends showed positive associations of physical activity with overall MHQ and each MHQ subcategory across all age groups. Larger estimated effects were observed for young and middle-aged adults as well as those 85+ years of age. Inspection of the estimated effects on the six MHQ subcategories suggest

that younger age groups may experience greater benefits from physical activity for Core Cognition and Adaptability and Resilience, as compared to the other older age groups. All ATCs and standard errors can be found in Supplementary Materials Table 2.

**Sensitivity analyses.** Our sensitivity analyses (see Table 1) demonstrated convergence of the estimated effects of physical activity on mental health across each of the alternative statistical techniques employed when compared to the main GBM results.

### Discussion

The purpose of the present study was to use propensity score weighting to estimate the association of physical activity engagement with a comprehensive indicator of mental health and its subcategories, and whether these effects may differ across age cohorts. Our findings revealed a significant small association of self-reported physical activity on overall mental health and well-being. Sensitivity analyses revealed this association was robust after adjusting for covariates using several different statistical methods. These findings are consistent with the existing evidence that has demonstrated beneficial associations between physical activity and various psychological outcomes including severe mental disorders (Vancampfort et al., 2017), cognitive function (Bidzan-Bluma & Lipowska, 2018; Carvalho et al., 2014), emotional skills (Laborde et al., 2016), resilience (Shanahan et al., 2022), and quality of life (Marquez et al., 2020; Schuch, Vancampfort, Rosenbaum, et al., 2016). Our results also align with previous work that has shown favorable effects of physical activity on indicators of mental health are consistent across the adult lifespan (Cunningham et al., 2020; de Oliveira et al., 2019). Taken together, these findings further underscore the importance of promoting physical activity, especially in those who are sedentary, to improve population mental health and well-being.

Findings from the present study also contribute to the body of literature investigating associations between physical activity and mental health through examining specific subcategories of mental health and well-being. Our results showed robust and consistent beneficial associations of physical activity on each of the six subcategories of the MHQ, although it should be noted that physical activity appears to have a significantly stronger relationship with Mind-Body Connection compared to Social Self as evidenced by non-overlapping 95% CIs. The Mind-Body Connection subcategory, in which the largest associations were observed for physical activity, contains items assessing aspects of well-being with benefits consistently related to physical activity, such as pain (Rice et al., 2019; Shiri & Falah-Hassani, 2017), sleep (Kredlow et al., 2015; Lederman et al., 2019), appetite regulation (Beaulieu et al., 2016, 2018), and fatigue (Bower, 2014; Pilutti et al., 2013). Conversely, the smallest association was shown for Social Self, which includes aspects of well-being with less established and robust associations to physical activity such as empathy, communication skills and relationship building (Pels & Kleinert, 2016; Shima et al., 2021). Although physical activity is a social pursuit for many, one potential explanation for a smaller effect of physical activity on Social Self is that some individuals prefer to engage in independent activities such as running or cycling alone, thus avoiding the potential social benefits associated with more group-oriented activities such as team sports and exercise classes. Nevertheless, these findings suggest that physical activity may confer benefits for all aspects of mental health and well-being, with small differences favoring psychophysiological over social aspects.

This study also addressed a knowledge gap regarding potential differential effects of physical activity on specific aspects of mental health and well-being across the adult lifespan. Evidence indicated that young and middle-aged adults may experience greater benefits for their

overall mental health from physical activity engagement in comparison to older adults. It should be noted that adults 85+ years of age appear to be an exception; however, this group also had the largest confidence interval likely due to a relatively smaller sample. As average levels of physical activity tend to be higher among young and middle-aged adults than older adults (Van Der Zee et al., 2019), a more sedentary lifestyle may also be especially indicative of impairment in younger cohorts. Core Cognition, Drive and Motivation, and Adaptability and Resilience followed the same trend as overall MHQ scores, and thus these specific aspects of mental health may benefit more by younger people adopting a more active lifestyle, whereas the other subcategories seem to demonstrate relatively consistent benefits from physical activity across the adult lifespan.

As the body of literature examining the effect of physical activity on mental health continues to grow, it is important that researchers adopt statistical best practices that can reduce bias and strengthen causal inferences when randomized controlled trials are unfeasible or unavailable. Matching and weighting techniques, though well established in the statistical literature, have received little attention in the fields of exercise psychology and behavioral medicine to date. The advantage of utilizing propensity scores over controlling for covariates in a traditional multivariable linear regression model is that the propensity model can deal with non-linear relationships between the covariates and outcome, as well as higher order interactions, and the GBM can handle these interaction terms non-parametrically when estimating the propensity score. As various propensity score estimation methods may perform differently based on assumptions and approaches to handle missing data (Cham & West, 2016; Coffman et al., 2020; Setodji et al., 2017), we utilized several combinations of methods in our sensitivity analyses. Our sensitivity analysis revealed minimal deviance in the estimated effects of physical activity on

overall MHQ scores across the various covariate adjustment and missing data procedures that were implemented – effect sizes were equivalent ranging from an SMD of 0.25 to 0.26. Using these various propensity score estimation techniques helped to improve our confidence that the estimated ATC was not biased due to misspecification of the propensity model. Though strong causal inferences are not indicated by cross-sectional observations, our results converge with existing intervention studies. For example, previous meta-analyses of randomized controlled trials on exercise and depression found pooled effect sizes ranging from 0.62 to 0.98 (Cooney et al., 2013; Josefsson et al., 2014; Schuch, Vancampfort, Richards, et al., 2016). Compared to previous intervention research however, our findings may underestimate the true effect of physical activity due to issues related with self-report (e.g., overestimating one's physical activity). Nevertheless, the present findings support and extend the existing literature on the benefits of physical activity engagement across various aspects of mental health.

Despite the strengths, there are several limitations with the current study. Firstly, unlike randomization, propensity score weighting does not balance unobserved covariates across treatment groups (Joffe & Rosenbaum, 1999). An unbiased treatment effect that can be interpreted as causal assumes that all potential confounders are observed, which is unlikely to be the case in any observational study concerning complex constructs and behaviors, such as mental health and physical activity engagement. The covariates included in this analysis were restricted by what was included in the GMP questionnaire, and adjusting for a partial set of confounders may reduce bias, but it is unknown to what extent. Second, the MHQ and its subcategories have yet, to our knowledge, been validated in an independent sample. It would be interesting, for example, to investigate whether the MHQ and its subcategories predict the onset or course of distinct mental disorders. Third, physical activity was self-reported, which can introduce recall

errors – particularly among older adults who are more prone to cognitive decline – and social desirability effects (Sallis & Saelens, 2000). However, researchers need to balance feasibility with practicality and therefore using a self-reported measure of physical activity may be best suited for data collection with a sample of this size and geographic spread. Lastly, the GMP has used convenience sampling to recruit participants, targeted towards individuals who used mental health-related search terms in Google and Facebook. Although the present sample includes individuals from over 200 countries and territories, it may not truly be globally representative, as it would likely have overlooked individuals living in regions with limited to no internet access.

### **Conclusion**

This cross-sectional study estimated the association of physical activity on several aspects of mental health among a large global sample of adults using propensity score weights estimated with generalized boosted modeling. Our results demonstrate a significant small effect of self-reported physical activity on a comprehensive measure of overall mental health and well-being, in addition to similar benefits across several specific subcategories of mental health. The strongest associations between physical activity and mental health appear to occur during the early and middle-aged adult life stages, with effects becoming weaker into old age. These findings further support the growing body of evidence promoting the benefits of physical activity on various aspects of mental health and well-being among the population, while demonstrating a useful statistical method for balancing covariates in observational data.

## References

- Aylett, E., Small, N., & Bower, P. (2018). Exercise in the treatment of clinical anxiety in general practice—a systematic review and meta-analysis. *BMC Health Services Research*, 18, 1–18.
- Beaulieu, K., Hopkins, M., Blundell, J., & Finlayson, G. (2016). Does habitual physical activity increase the sensitivity of the appetite control system? A systematic review. *Sports Medicine*, 46, 1897–1919.
- Beaulieu, K., Hopkins, M., Blundell, J., & Finlayson, G. (2018). Homeostatic and non-homeostatic appetite control along the spectrum of physical activity levels: An updated perspective. *Physiology & Behavior*, 192, 23–29.
- Bidzan-Bluma, I., & Lipowska, M. (2018). Physical Activity and Cognitive Functioning of Children: A Systematic Review. *International Journal of Environmental Research and Public Health*, 15(4), 800. <https://doi.org/10.3390/ijerph15040800>
- Borsboom, D., Cramer, A. O., Schmittmann, V. D., Epskamp, S., & Waldorp, L. J. (2011). The small world of psychopathology. *PloS One*, 6(11), e27407.
- Boschloo, L., Bekhuis, E., Weitz, E. S., Reijnders, M., DeRubeis, R. J., Dimidjian, S., Dunner, D. L., Dunlop, B. W., Hegerl, U., Hollon, S. D., & others. (2019). The symptom-specific efficacy of antidepressant medication vs. Cognitive behavioral therapy in the treatment of depression: Results from an individual patient data meta-analysis. *World Psychiatry*, 18(2), 183–191.
- Bower, J. E. (2014). Cancer-related fatigue—Mechanisms, risk factors, and treatments. *Nature Reviews Clinical Oncology*, 11(10), 597–609.
- Burgette, J. M., Preisser, J. S., & Rozier, R. G. (2016). Propensity score weighting: An application to an Early Head Start dental study. *Journal of Public Health Dentistry*, 76(1), 17–29.
- Carvalho, A., Rea, I. M., Parimon, T., & Cusack, B. J. (2014). Physical activity and cognitive function in individuals over 60 years of age: A systematic review. *Clinical Interventions in Aging*, 661–682.

- 344 Cham, H., & West, S. G. (2016). Propensity score analysis with missing data. *Psychological Methods*,  
345 21(3), 427.
- 346 Chekroud, S. R., Gueorguieva, R., Zheutlin, A. B., Paulus, M., Krumholz, H. M., Krystal, J. H., & Chekroud,  
347 A. M. (2018). Association between physical exercise and mental health in 1· 2 million individuals  
348 in the USA between 2011 and 2015: A cross-sectional study. *The Lancet Psychiatry*, 5(9), 739–  
349 746.
- 350 Coffman, D. L., Zhou, J., & Cai, X. (2020). Comparison of methods for handling covariate missingness in  
351 propensity score estimation with a binary exposure. *BMC Medical Research Methodology*, 20(1),  
352 1–14.
- 353 Cooney, G. M., Dwan, K., Greig, C. A., Lawlor, D. A., Rimer, J., Waugh, F. R., McMurdo, M., & Mead, G. E.  
354 (2013). Exercise for depression. *Cochrane Database of Systematic Reviews*, 9.
- 355 Cunningham, C., O'Sullivan, R., Caserotti, P., & Tully, M. A. (2020). Consequences of physical inactivity in  
356 older adults: A systematic review of reviews and meta-analyses. *Scandinavian Journal of*  
357 *Medicine & Science in Sports*, 30(5), 816–827.
- 358 de Oliveira, L. da S. S. C. B., Souza, E. C., Rodrigues, R. A. S., Fett, C. A., & Piva, A. B. (2019). The effects of  
359 physical activity on anxiety, depression, and quality of life in elderly people living in the  
360 community. *Trends in Psychiatry and Psychotherapy*, 41, 36–42.
- 361 Ekkekakis, P. (2020). Why Is exercise underutilized in clinical practice despite evidence it is effective?  
362 Lessons in pragmatism from the inclusion of exercise in guidelines for the treatment of  
363 depression in the British National Health Service. *Kinesiology Review*, 10(1), 29–50.
- 364 Fessler, J., Godement, M., Pirracchio, R., Marandon, J.-Y., Thes, J., Sage, E., Roux, A., Parquin, F., Cerf, C.,  
365 Fischler, M., & others. (2019). Inhaled nitric oxide dependency at the end of double-lung  
366 transplantation: A boosted propensity score cohort analysis. *Transplant International*, 32(3),  
367 244–256.



- 368 Fried, E. (2017). Moving forward: How depression heterogeneity hinders progress in treatment and  
369 research. *Expert Review of Neurotherapeutics*, 17(5), 423–425.
- 370 Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of*  
371 *Statistics*, 1189–1232.
- 372 Funk, M. J., Westreich, D., Wiesen, C., Stürmer, T., Brookhart, M. A., & Davidian, M. (2011). Doubly  
373 robust estimation of causal effects. *American Journal of Epidemiology*, 173(7), 761–767.
- 374 Gordon, B. R., McDowell, C. P., Hallgren, M., Meyer, J. D., Lyons, M., & Herring, M. P. (2018). Association  
375 of efficacy of resistance exercise training with depressive symptoms: Meta-analysis and meta-  
376 regression analysis of randomized clinical trials. *JAMA Psychiatry*, 75(6), 566–576.
- 377 Greifer, N. (2022). *WeightIt: Weighting for Covariate Balance in Observational Studies*. [https://CRAN.R-](https://CRAN.R-project.org/package=WeightIt)  
378 [project.org/package=WeightIt](https://CRAN.R-project.org/package=WeightIt)
- 379 Iniesta, R., Malki, K., Maier, W., Rietschel, M., Mors, O., Hauser, J., Henigsberg, N., Dernovsek, M. Z.,  
380 Souery, D., Stahl, D., & others. (2016). Combining clinical variables to optimize prediction of  
381 antidepressant treatment outcomes. *Journal of Psychiatric Research*, 78, 94–102.
- 382 Joffe, M. M., & Rosenbaum, P. R. (1999). Invited commentary: Propensity scores. *American Journal of*  
383 *Epidemiology*, 150(4), 327–333.
- 384 Josefsson, T., Lindwall, M., & Archer, T. (2014). Physical exercise intervention in depressive disorders:  
385 Meta-analysis and systematic review. *Scandinavian Journal of Medicine & Science in Sports*,  
386 24(2), 259–272.
- 387 Kendrick, T., Pilling, S., Mavranezouli, I., Megnin-Viggars, O., Ruane, C., Eadon, H., & Kapur, N. (2022).  
388 Management of depression in adults: Summary of updated NICE guidance. *BMJ*, 378, o1557.  
389 <https://doi.org/10.1136/bmj.o1557>
- 390 Kredlow, M. A., Capozzoli, M. C., Hearon, B. A., Calkins, A. W., & Otto, M. W. (2015). The effects of  
391 physical activity on sleep: A meta-analytic review. *Journal of Behavioral Medicine*, 38, 427–449.

- 392 Laborde, S., Dosseville, F., & Allen, M. S. (2016). Emotional intelligence in sport and exercise: A  
393 systematic review. *Scandinavian Journal of Medicine & Science in Sports*, 26(8), 862–874.
- 394 Lanza, S. T., Moore, J. E., & Butera, N. M. (2013). Drawing causal inferences using propensity scores: A  
395 practical guide for community psychologists. *American Journal of Community Psychology*, 52,  
396 380–392.
- 397 Lederman, O., Ward, P. B., Firth, J., Maloney, C., Carney, R., Vancampfort, D., Stubbs, B., Kalucy, M., &  
398 Rosenbaum, S. (2019). Does exercise improve sleep quality in individuals with mental illness? A  
399 systematic review and meta-analysis. *Journal of Psychiatric Research*, 109, 96–106.
- 400 Lee, B. K., Lessler, J., & Stuart, E. A. (2010). Improving propensity score weighting using machine  
401 learning. *Statistics in Medicine*, 29(3), 337–346.
- 402 Lee, B. K., Lessler, J., & Stuart, E. A. (2011). Weight trimming and propensity score weighting. *PloS One*,  
403 6(3), e18174.
- 404 Malhi, G. S., Bassett, D., Boyce, P., Bryant, R., Fitzgerald, P. B., Fritz, K., Hopwood, M., Lyndon, B.,  
405 Mulder, R., Murray, G., Porter, R., & Singh, A. B. (2015). Royal Australian and New Zealand  
406 College of Psychiatrists clinical practice guidelines for mood disorders. *Australian and New*  
407 *Zealand Journal of Psychiatry*, 49, 1087–1206. <https://doi.org/10.1177/0004867415617657>
- 408 Marquez, D. X., Aguiñaga, S., Vázquez, P. M., Conroy, D. E., Erickson, K. I., Hillman, C., Stillman, C. M.,  
409 Ballard, R. M., Sheppard, B. B., Petruzzello, S. J., & others. (2020). A systematic review of  
410 physical activity and quality of life and well-being. *Translational Behavioral Medicine*, 10(5),  
411 1098–1109.
- 412 McCaffrey, D. F., Ridgeway, G., & Morral, A. R. (2004). Propensity score estimation with boosted  
413 regression for evaluating causal effects in observational studies. *Psychological Methods*, 9(4),  
414 403.

- 415 McDowell, C. P., Dishman, R. K., Gordon, B. R., & Herring, M. P. (2019). Physical activity and anxiety: A  
416 systematic review and meta-analysis of prospective cohort studies. *American Journal of*  
417 *Preventive Medicine*, 57(4), 545–556.
- 418 Murri, M. B., Ekkekakis, P., Menchetti, M., Neviani, F., Trevisani, F., Tedeschi, S., Latessa, P. M., Nerozzi,  
419 E., Ermini, G., Zocchi, D., & others. (2018). Physical exercise for late-life depression: Effects on  
420 symptom dimensions and time course. *Journal of Affective Disorders*, 230, 65–70.
- 421 Newson, J. J., Pastukh, V., & Thiagarajan, T. C. (2021). Poor separation of clinical symptom profiles by  
422 DSM-5 disorder criteria. *Frontiers in Psychiatry*, 12, 775762.
- 423 Newson, J. J., Pastukh, V., & Thiagarajan, T. C. (2022). Assessment of Population Well-being With the  
424 Mental Health Quotient: Validation Study. *JMIR Mental Health*, 9(4), e34105.
- 425 Newson, J. J., & Thiagarajan, T. C. (2020). Assessment of population well-being with the mental health  
426 quotient (MHQ): Development and usability study. *JMIR Mental Health*, 7(7), e17935.
- 427 Pearce, M., Garcia, L., Abbas, A., Strain, T., Schuch, F. B., Golubic, R., Kelly, P., Khan, S., Utukuri, M., Laird,  
428 Y., & others. (2022). Association between physical activity and risk of depression: A systematic  
429 review and meta-analysis. *JAMA Psychiatry*.
- 430 Pels, F., & Kleinert, J. (2016). Loneliness and physical activity: A systematic review. *International Review*  
431 *of Sport and Exercise Psychology*, 9(1), 231–260.
- 432 Pilutti, L. A., Greenlee, T. A., Motl, R. W., Nickrent, M. S., & Petruzzello, S. J. (2013). Effects of exercise  
433 training on fatigue in multiple sclerosis: A meta-analysis. *Psychosomatic Medicine*, 75(6), 575–  
434 580. <https://doi.org/10.1097/PSY.0b013e31829b4525>
- 435 Pollack, C. E., Griffin, B. A., & Lynch, J. (2010). Housing affordability and health among homeowners and  
436 renters. *American Journal of Preventive Medicine*, 39(6), 515–521.
- 437 R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for  
438 Statistical Computing. <https://www.R-project.org/>

- 439 Rice, D., Nijs, J., Kosek, E., Wideman, T., Hasenbring, M. I., Koltyn, K., Graven-Nielsen, T., & Polli, A.  
440 (2019). Exercise-induced hypoalgesia in pain-free and chronic pain populations: State of the art  
441 and future directions. *The Journal of Pain*, 20(11), 1249–1266.
- 442 Ridgeway, G., McCaffrey, D. F., Morral, A. R., Cefalu, M., Burgette, L. F., Pane, J. D., & Griffin, B. A.  
443 (2022). *Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the R TWANG*  
444 *package*. Rand Santa Monica, Calif.
- 445 Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies  
446 for causal effects. *Biometrika*, 70(1), 41–55.
- 447 Sallis, J. F., & Saelens, B. E. (2000). Assessment of physical activity by self-report: Status, limitations, and  
448 future directions. *Research Quarterly for Exercise and Sport*, 71(sup2), 1–14.
- 449 Schuch, F. B., Vancampfort, D., Richards, J., Rosenbaum, S., Ward, P. B., & Stubbs, B. (2016). Exercise as a  
450 treatment for depression: A meta-analysis adjusting for publication bias. *Journal of Psychiatric*  
451 *Research*, 77, 42–51.
- 452 Schuch, F. B., Vancampfort, D., Rosenbaum, S., Richards, J., Ward, P. B., & Stubbs, B. (2016). Exercise  
453 improves physical and psychological quality of life in people with depression: A meta-analysis  
454 including the evaluation of control group response. *Psychiatry Research*, 241, 47–54.
- 455 Setodji, C. M., McCaffrey, D. F., Burgette, L. F., Almirall, D., & Griffin, B. A. (2017). The right tool for the  
456 job: Choosing between covariate balancing and generalized boosted model propensity scores.  
457 *Epidemiology (Cambridge, Mass.)*, 28(6), 802.
- 458 Shanahan, L., Steinhoff, A., Bechtiger, L., Murray, A. L., Nivette, A., Hepp, U., Ribeaud, D., & Eisner, M.  
459 (2022). Emotional distress in young adults during the COVID-19 pandemic: Evidence of risk and  
460 resilience from a longitudinal cohort study. *Psychological Medicine*, 52(5), 824–833.

- 461 Shima, T., Jesmin, S., Nakao, H., Tai, K., Shimofure, T., Arai, Y., Kiyama, K., & Onizawa, Y. (2021).  
462 Association between self-reported empathy and level of physical activity in healthy young  
463 adults. *The Journal of Physical Fitness and Sports Medicine*, 10(1), 45–49.
- 464 Shin, H.-J., Son, S.-Y., Wang, B., Roh, C. K., Hur, H., & Han, S.-U. (2021). Long-term comparison of robotic  
465 and laparoscopic gastrectomy for gastric cancer: A propensity score-weighted analysis of 2084  
466 consecutive patients. *Annals of Surgery*, 274(1), 128–137.
- 467 Shiri, R., & Falah-Hassani, K. (2017). Does leisure time physical activity protect against low back pain?  
468 Systematic review and meta-analysis of 36 prospective cohort studies. *British Journal of Sports*  
469 *Medicine*, 51(19), 1410–1418.
- 470 Stubbs, B., Vancampfort, D., Hallgren, M., Firth, J., Veronese, N., Solmi, M., Brand, S., Cordes, J.,  
471 Malchow, B., Gerber, M., Schmitt, A., Correll, C. U., Hert, M. D., Gaughran, F., Schneider, F.,  
472 Kinnaifick, F., Falkai, P., Möller, H.-J., & Kahl, K. G. (2018). EPA guidance on physical activity as a  
473 treatment for severe mental illness: A meta-review of the evidence and Position Statement from  
474 the European Psychiatric Association (EPA), supported by the International Organization of  
475 Physical Therapists in Mental Health (IOPTMH). *European Psychiatry*, 54, 124–144.  
476 <https://doi.org/10.1016/j.eurpsy.2018.07.004>
- 477 Tu, C. (2019). Comparison of various machine learning algorithms for estimating generalized propensity  
478 score. *Journal of Statistical Computation and Simulation*, 89(4), 708–719.
- 479 Van Der Zee, M. D., Van Der Mee, D., Bartels, M., & De Geus, E. J. (2019). Tracking of voluntary exercise  
480 behaviour over the lifespan. *International Journal of Behavioral Nutrition and Physical Activity*,  
481 16(1), 1–11.
- 482 Vancampfort, D., Firth, J., Schuch, F. B., Rosenbaum, S., Mugisha, J., Hallgren, M., Probst, M., Ward, P. B.,  
483 Gaughran, F., De Hert, M., & others. (2017). Sedentary behavior and physical activity levels in

people with schizophrenia, bipolar disorder and major depressive disorder: A global systematic review and meta-analysis. *World Psychiatry*, 16(3), 308–315.

Von Elm, E., Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: Guidelines for reporting observational studies. *The Lancet*, 370(9596), 1453–1457.

508 **Tables and Figures**

509

510 **Table 1. Sensitivity analysis. Reported ATC effects and standard errors on MHQ**

		Doubly		Doubly		Doubly
		Robust		Robust		Robust MI
MHQ	GBM	GBM	MI + GBM	MI + GBM	MI + CBPS	+ CBPS
ATC	17.86	17.74	17.75	17.77	18.15	17.87
SE	1.43	0.99	0.28	0.23	0.28	0.22
SMD	0.25	0.25	0.25	0.25	0.26	0.25

511

512

513

514

515

516

517

518

519

520

521

522

523

**Figure 1.** ATC’s of physical activity on overall MHQ and subcategories scores across age groups. Error bars represent 95% CIs.

