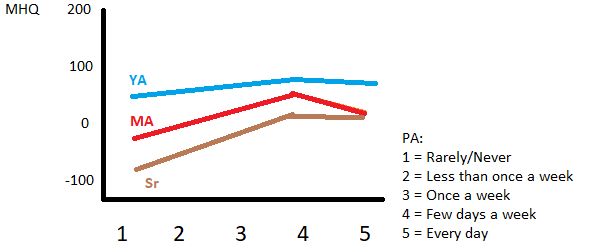
1. Data wrangling
   1. Subset outcomes, treatment variable, and covariates
   2. Recode and consolidate
   3. Drop ethnicity and genderdiff due to missing data (mice errored)
2. Treat PA as ordinal, nominal, or interval?
   1. Test whether interval is ok
      1. LR-test
      2. Wald-test
      3. Sheaf coefficients

1. Multiple Imputation of Missing Data
   1. Create prediction matrix, zero out ID and denote country as class variable (-2)
   2. Set imputation methods (pmm default, logreg and polyreg for factors)
   3. Run imputation
   4. Check imputed data
2. Covariate Balancing
   1. Generate balancing weights for covariates to estimate treatment effect of PA (multi-category)
      1. Compute a weighted dataset estimating ATT with focal = “Rarely/Never”
      2. Compute a weighted dataset estimating ATE (no focal variable) (nope)
      3. Compute a weighted dataset estimating ATE with binary treatment (Meet PA guidelines yes/no) (Should we do this? Need to recode PA to binary and re impute missing data)
   2. Check balance
      1. Love plot
3. Estimation of Treatment Effect
   1. Formulate a multi-level (nested by country) survey weighted GLM, and include PA\*age (uncategorized age factors)
      1. Pool estimates and print results for 6 outcomes: MHQ, Cognition, Adaptability and Resilience, Drive and Motivation, Mood and Outlook, Social Self, Mind-Body Connection



* + - * Separate analysis for mhseeking = Yes? (N=50606 ~ 14% of sample)
      * Sensitivity analysis with unweighted sample?
      * Separate analysis for male:female
      * Sex differences (18-24 girls high PA = low MH? Due to BI/ED)
        1. 4 categories of age
        2. 18-24 = early adult
        3. 25-34 = young adult
        4. 35-44, 45-54, 55-64 = middle adult
        5. 65-74, 75-84, 85+ = senior

Propensity scores were then winsorised at a 99% level to minimise the impact of excessive values”

“Non-parametric twosample Wilcoxon rank-sum tests were used to assess for differences in mental health burden between these matched groups (figure 1). Since a previous diagnosis of depression could have an extremely strong association with current mental health burden, we did separate matched sample analyses for individuals who had been diagnosed with depression in the past and those with no previous diagnosis of depression. Finally, to ensure that the findings were not an artifact of covariate adjustment, we did sensitivity analyses without matching procedures”

We formally analysed the effects of exercise duration and frequency using a generalised additive model to allow us to observe non-linear relationships with mental health burden. The model used penalised cubic regression splines for exercise duration and for frequency, with parametric regressors to control for the full covariate set. We then plotted the fitted smoothed coefficients for duration and frequency, with 95% CIs