Supplementary material

S1. Statistical analysis

We aim to explore the three types of regression models as each one makes different assumptions and treats the outcome in a different way. MTBQ scores are likely to be skewed but given the relatively large sample size, any non-normality of residuals in the linear regression is unlikely to affect our inference [1]. For the ordinal logistic regression, we will assess the parallel regression assumption using a Brant test. Treating the outcome as continuous should give greater power to detect associations. However, if we find strong evidence the assumptions of this model are not held (e.g., due to heteroscedasticity of residuals) the results from either the ordinal or logistic regression will be presented as the primary results. Our imputed analyses will use multiple imputation with the chained equations approach. We will use fractional polynomials to assess non-linear associations between numeric variables and the outcome.

1 Schmidt, A.F. and C. Finan, *Linear regression and the normality assumption.* J Clin Epidemiol, 2018. **98**: p. 146-151.