Preregistration Template for Coordinated Data Analysis

Part 1: Dataset Inclusion and Exclusion Criteria

a. How were datasets identified and selected? For example, did the researchers conduct a systematic search of data repositories? If so, which repositories and search terms were included in the search? Did the search process include strategies for locating grey datasets (i.e., datasets that are not located in data repositories)? If the researchers did not conduct a systematic search of relevant datasets, this should be noted.

b. What are the minimum inclusion and exclusion criteria for including a dataset in the coordinated data analysis?

c. Which specific datasets meet the criteria outlined in 1b and will be included in the coordinated data analysis?

d. If additional datasets, waves, or cohorts that are not named in 1c are identified that meet the criteria outlined in 1b, will they be added to the project? If yes, what is the latest stage at which additional data will be added?

e. If datasets, waves, or cohorts identified in 1c are later found not to meet the criteria outlined in 1b, will they be dropped from the project? If yes, what is the latest stage at which data will be dropped?

Part 2: Variable Harmonization

a. For each variable in the study, please outline the degree of flexibility that you will allow in its operational definition. E.g., How much variation in the scale(s) used and/or the response options given will you allow?

b. For each variable in the study, please outline your harmonization plan including any data transformations. If the exact variable types and response options are not yet known or at not

known with confidence, please provide contingency plans for each possible variable type and set of response options.

Part 3: Model Harmonization (i.e., Individual Study Analyses)

a. What is the optimal statistical model you will use to evaluate each hypothesis within the individual datasets?

b. What is the minimum viable model you will use to evaluate each hypothesis within the individual datasets?

c. (optional) If the models described in 3a and 3b do not reflect the full range of models that will be evaluated, outline an organizational chart of possible models and corresponding data requirements.

Part 4: Results Synthesis and Reporting (i.e., Meta-Analyses)

a. How will parameter estimates be summarized to evaluate each hypothesis (e.g., individual study estimates only, mean weighted effect sizes, random effects or fixed effects meta-analysis)? If results will be summarized, will individual study results also be presented? How will between-study heterogeneity be evaluated, if at all?

b. Will parameter estimates from non-identical models be combined? If yes, how? If no, how will hypotheses be evaluated?

c. (optional) How will individual study results and synthesized results be plotted and/or visualized?

Part 5: Supporting Documents (optional)

a. Instructions for external study analysts: If more than one analyst is involved in the coordinated data analysis, general instructions should be provided to analysts regarding how to use other supporting documents. These instructions can be attached to this pre-registration.

b. Dataset construction document: Specific information about how to prepare individual datasets can be attached to this pre-registration. E.g., How should data be structured (wide versus long)? How should variables be coded (e.g., what items should be included in composites) or transformed (e.g., dummy coding, reverse coding, standardizations)?

c. Statistical code: To ensure consistency across datasets, a single set of statistical code should be applied to each dataset. Contingencies based on the availability of specific variables and/or data types can be built into the statistical code and should be automated whenever possible. This statistical code, and additional statistical code for synthesized results across datasets, can be attached to this pre-registration.