A Guide for Protocols in Observational Studies

Mike Baiocchi, Jonathan Pipping, Andrea Schneider, Ashil Srivastava and Dylan Small

Contents

1   Title and Abstract (S1)

1.1   Goal of Section

The title and abstract provide a concise summary of the study for your readers. A clear title ensures the purpose of the study is clear and that the paper is easily searchable. The abstract provides a structured summary of the research process. Together, they ensure the transparency and accessibility of the study.

1.2   Steps:

- **Title:** Choose a title that describes the study accurately and succinctly. Make sure to include a commonly-used term that describes the study type (e.g., "A Matching Study on the Effects of High School Football on Cognition Late in Life.").

- **Abstract:** Summarize the study in a structured format, typically including Background, Methods, Results, and Conclusion. Ensure the abstract is informative and balanced, accurately reflecting what was done and found.

2   Introduction

2.1   Goal of Section

The introduction sets the stage for the study by explaining background information, how it motivates your research question, and your rationale in defining the study's methods and objectives. It provides the reader with context and highlights the importance and relevance of your research.

2.2   Steps:

- **Literature Review (S2):** Summarize existing research relevant to the study topic. Make sure to cite sources and be thorough; this will help the reader understand why your study is important and how it fits into existing work.

- **Data and Sources (S8):** Identify the datasets used and sources or providers. Be as specific as possible. Additionally, describe how much access you had to the data when writing the protocol.

- **Setting (S5):** Describe when and where data collection occurred. Was it in a specific location or time period? How were subjects selected? Be as descriptive as possible.

- **Motivation (S2):** Explain why this study is important and what knowledge gap it fills. This should follow naturally from the literature review.

3   Research Question

3.1   Goal of Section

This section clearly defines the research question. The section also elucidates the population or setting to which the research results are intended to apply.

3.2   Steps

- **Hypotheses and Objectives (S3):** Clearly state the research question, hypotheses, and objectives. This will help guide your study design and ensure transparency in selected methods.

- **Generalization Target (S21):** Define the population or setting to which potential findings would (and wouldn't) apply.

4 Inclusion and Exclusion Criteria

4.1 Goal of Section

This section defines which participants or data points will be included or excluded, ensuring transparency, clarity and reproducibility. This also serves to prevent selection bias in the study.

4.2 Steps:

- **Inclusion Criteria (S6):** Define the selection criteria for study participants or data points. This should be specific and clear.

- **Exclusion Criteria (S13):** State reasons for exclusion of certain participants, such as missing or un-usable data. Be specific about how many were excluded and why.

- **Data Counts (S10):** Report the number of participants or data points at each stage of the analysis. This will help readers understand the study's progression.

- **Flowchart:** Visual data counts by including a flowchart summarizing the selection and exclusion processes at each step of the analysis. Include the number of participants or data points at each stage and an explanation for exclusions.

- **Generalization Considerations (S21):** Which subpopulations are included or excluded? Discuss how these criteria affect the study's generalizability and applicability.

5 Study Outcomes (S7, S15)

5.1 Goal of Section

Defines the primary and secondary outcomes that the study aims to analyze. This ensure that the study's objectives are clear to the reader and that the study is reproducible.

5.2 Steps:

- **Primary Outcome:** Define the main variable being studied as the outcome. If you plan to modify a variable in the data set to then use as the outcome variable be sure describe any adjustments (e.g., re-coding of a variable, thresholding to convert a continuous variable to a binary).

- **Connection to Study:** Explain how the outcome connects to the research question, hypotheses, and objectives.

- **Marginal Distributions:** Display a histogram or box plot of the outcome variable.

- **Secondary Outcomes:** Additional variables planned to be analyzed separately. For each, follow the same steps as for the primary outcome.

6 Study Design (S4, S12)

6.1 Goal of Section

Explains how the study is structured and what methods are used. This ensures that the study is well-defined and reproducible.

6.2 Steps:

- **Primary Design Feature (S6):** Define key methodological elements and their purpose (e.g., matching to reduce covariate imbalance, instrumental varibles to address bias from unobserved covariates). What is the intended estimand? How is estimation supported by the design?

- **Potential Observed Confounders (S7):** Identify which covariates will be controlled for, or adjusted for, in order to reduce bias in estimation. Note that this list may overlap with the list in "Effect Modifiers."

- **Potential Unobserved Confounders (S7):** Describe any covariates that are either not measured, or not sufficiently measured (e.g., only a proxy is available, or measured with differential error), which have the potential to bias estimation.

- **Effect Modifiers (S7):** Identify any covariates believed to be effect modifiers. Note that this list may overlap with the list in "Potential Confounders."

- **Mediators (S7):** If the study is attempting decompose a treatment effect into sub-causes (e.g., direct vs indirect effect) then identify any variable believed to be on the causal pathways. In most studies attempting to analyze such decompositions, it will be useful to use a directed acyclic graph to communicate beliefs about the causal pathways and the interrelatedness of the full set of covariates with the outcome.

- **Precision Variables (S7):** Identify and describe non-outcome variables. How are they related to the outcome? How are confounders addressed by the study design?

- **Variable Selection (S7):** Describe how variables were selected for inclusion in the analysis. Were data-driven variable selection procedures used? In some settings, it may make sense to use a directed acyclic graph to communicate how variables were selected for consideration.

- **Generalizability (S21):** How does the study design and confounder control affect the study's generalizability?

- **Handling Missing/Corrupted Data (S12, S14):** Explain any imputation, exclusion, or other methods used to address missing or corrupted data.

- **Bias Control from Observed Covariates(S9):** Describe how possible bias is addressed. Examples include selection bias, measurement error, and confounding.

- **Methods for Addressing or Evaluating Bias from Unobserved Covariates:** Describe methods like negative controls (a.k.a., "known null analysis"), multiple analyses with differentiated sources biases (e.g., triangulation), or evidence factors.

- **Implementation (S11, S12):** Describe how the study design will be implemented (e.g., software packages). This may include providing code.

- **Covariate Balance and Distributions (S14):** Present visualizations demonstrating the balance of covariates across treatment groups (e.g., before and after matching). Provide other means of verifying successful implementation of the study design.

- **Secondary Features (S6):** Describe any additional other methodological elements and their purpose. Follow the same steps as for primary features.

7 Study Analysis Plan (S12)

7.1 Goal of Section

Describes how processed data will be analyzed and interpreted. Ensures transparency with the reader and reproducibility of the study. This section requires careful statistical reasoning and description of procedures; in many situations, having a (bio)statistician or similarly trained researcher involved in developing this section will be beneficial.

7.2 Steps:

- **Primary Analysis:** Define and describe statistical tests and associated thresholds used to test the hypothesis of interest. What is the role of the primary outcome variable in this analysis?

- **Connection to Hypotheses:** How do the proposed statistical tests connect to the research question and hypotheses?

- **Multiple Testing:** If you have multiple hypotheses, how will you address multiple testing? (e.g., to ensure familywise error rate control or false discovery rate control)

- **Inferential Methods (S12):** Define hypothesis tests and confidence intervals. What sort of answer will these tests provide to the research question? What level of confidence is appropriate?

- **Challenges and Solutions (S19):** Describe any anticipated challenges that may arise during the analysis. How will these be addressed?

- **Empirical Summary (S14):** Describe summary statistics, tables, and visualizations that will be used to report findings. Consider producing "mocked up" versions of these summaries (e.g., using simulated data) to help readers understand anticipated reporting format.

- **Result Interpretation (S20):** Explain how testing results will be interpreted. Examples include: describing anticipated range of possible results, statistically significant vs non-significant results, what magnitude of effect sizes is meaningful to the literature. How will the results of this test be used to answer the research questions?

- **Generalizability (S21):** Discuss the applicability and generalizability of the study results.

8 Sensitivity Analysis (S12, S17)

8.1 Goal of Section

Tests the robustness of the study's conclusions to different assumptions and potential biases. Ensures that the study's results are reliable and not overly sensitive to specific assumptions. This may include evaluating how different definitions of exposure, outcome measurement variations, or sub-sample analyses might affect the results.

8.2 Steps:

- **Formal Sensitivity Methods:** Which methods are being used (e.g., gamma-sensitivity analysis). How does each method evaluate an inferential assumption (e.g., no bias from unmeasured covariates)?

- **Interpretation of Formal Sensitivity Analyses:** Explain how sensitivity analysis results will be interpreted. Examples include: describing anticipated range of possible results, how would different values of the sensitivity parameters be interpreted. How will the results of this test be used to answer the research questions?

- **Stability Analyses:** Describe any stability analyses (e.g., changes in variable definitions in order to assess changes in outcomes).

9 Simulation of Protocol

9.1 Goal of Section

As a means for carefully describing the steps described above in the protocol, this section suggests simulations as a way of developing implementation code and to more explicitly identify assumptions.

9.2 Steps:

- **Simulation of Data:** WORDS1.

- **Data Pipelining Code:** WORDS2.

- **Baseline Summary Code:** WORDS3.

- **Analyses Code:** WORDS4.

- **Results Summary Code:** WORDS5.

10 Amendments

10.1 Goal of Section

Documents any changes made across the study's life cycle. Ensures transparency and accounts for potential induced biases from post hoc changes.

10.2 Steps:

- **Clarifying Hypotheses:** Document any changes made to the study's hypotheses. Explain the rationale behind these changes and how they affect the study's objectives.

- **Change in Inclusion/Exclusion:** Describe any modifications to the inclusion or exclusion criteria. Provide reasons for these changes and discuss their potential impact on the study's generalizability.

- **Change in Data Processing:** Detail any changes made to data processing methods. Explain why these changes were necessary and how they affect the study results.

- **Change in Outcomes:** Note any changes to primary or secondary outcomes. Justify these changes and discuss their implications for the study's conclusions.

- **Change in Study Design:** Describe any alterations to the study design. Explain the reasons for these changes and their potential impact on the validity of the study.

- **Change in Study Analysis:** Document any changes to the analysis plan. Provide a rationale for these changes and discuss how they affect the interpretation of the study's results.

- **Change in Sensitivity Analysis:** Detail any modifications to the sensitivity analysis methods. Explain why these changes were made and their implications for the robustness of the study.

11 Funding (S22)

11.1 Goal of Section

Declares financial support and potential conflicts of interest. Ensures accountability and minimizes potential for undisclosed biases.

11.2 Steps:

- **Funding Source:** Identify the source of financial support for the study. Be specific and transparent.

- **Conflict of Interest:** Disclose any potential conflicts of interest that could influence the study results or interpretation. Be transparent and thorough.

- **Role of Funders:** Describe the role of funding sources in the study (e.g., study design, data collection, analysis). How did they contribute to the research process? How might this affect the study's results?

12   Future Steps (S16, S18)

12.1   Missing STROBE Points

These STROBE items pertain to reporting actual results and key findings once data analysis is complete. In this protocol, we plan to address S16 and S18 in a dedicated Results section after data collection and analysis.