ARNOLD VENTURES

Pre-Analysis Plan Guidance

For RCTs and QEDs[1]

Arnold Ventures is committed to the principles of research transparency and integrity. As a condition of any grant award, we ask awardees to pre-register empirical studies that involve statistical inference on the Open Science Framework website before the start of the intervention or data analysis.[2]

The key information to include in a pre-analysis plan depends on the particular study, and we ask researchers to touch base with us prior to drafting their plan to jointly identify appropriate items to include. In what follows, we provide a checklist of information that we typically look for in Randomized Controlled Trial (RCT) and Quasi-Experimental Design (QED) pre-analysis plans. In general, we expect pre-analysis plans to be no more than 6 pages in length.

When the research project is concluded, any report or article should conform as much as possible to the pre-registered design and analysis, with deviations being identified and explained. In many cases (particularly studies other than experiments), it may be infeasible to pre-register all modeling decisions that will be made along the way, although the basic modeling framework should be articulated clearly and justified, with caveats as to future decisions that depend on as-yet unknown circumstances. As further means of improving reliability, researchers should engage in thorough robustness and sensitivity testing as to their modeling assumptions and choices.

RCT PRE-ANALYSIS PLAN SUGGESTED OUTLINE

- **General Information**
  - Title of project
  - Researchers involved, including name, title, and institution
  - External partner institutions (e.g., implementing partner, if applicable)
- **Introduction**
  - Brief project summary (~1 paragraph), including description of intervention
- **Study design**
  - Identify 1-2 primary outcomes
    - *If >2 primary outcomes, describe the method for multiple comparisons adjustment*
    - *Note: in studies that collect outcomes over multiple time periods (e.g., two year college persistence and then three year college persistence), two year college persistence will be superseded by three year college persistence. In these cases, multiple comparison adjustment may not be required and we encourage you to discuss this with us.*
  - Identify secondary or exploratory outcomes (as applicable)
  - Definitions of outcomes (e.g., recidivism defined as number of arrests)
  - Data sources
  - Length of follow-up
  - Power analysis for primary outcomes, with minimum detectable effects translated into real world terms (e.g., 5 percentage point increase in degree receipt, $350 in earnings)
  - Analysis methods, preferably using intention-to-treat (ITT) for primary outcomes
  - Individual vs. cluster random assignment
    - *If cluster, the method for adjusting standard errors*
  - Covariate selection
  - Addressing missing data (as applicable)

---

1 This is an adapted version of the checklist in Chuang, E. and Wykstra, S. "A Guide to Pre-Analysis Plans," Innovations for Poverty Action. 2015.

2 Studies can be registered elsewhere (e.g., the AEA or ICPSR pre-registry), but OSF allows for open-ended pre-registration of study materials and pre-analysis plans, giving research teams greater flexibility to share information about study methods and adjustments made during the course of a study.

- Addressing data outliers (preferably no trimming)
  - *If trimming, the method for trimming and confirming the blinding of data analysts to which data are treatment vs. control*
- Accounting for stratification by site (as applicable)

- **Sample Selection**
  - Anticipated sample size and a discussion of inclusion/exclusion criteria
  - Clearly describing the treatment and control groups
  - Expected study enrollment timeline
  - Random assignment procedures, including at what point consent is received

QED PRE-ANALYSIS PLAN SUGGESTED OUTLINE

- **General Information**
  - Title of project
  - Researchers involved, including name, title, and institution
  - External partner institutions (e.g., implementing partner, if applicable)
- **Introduction**
  - Brief project summary (~1 paragraph), including description of intervention
- **Study design**
  - Identify 1-2 primary outcomes
    - *For each outcome, please specify the length of follow-up*
    - *If >2 primary outcomes, describe the method for multiple comparisons adjustment*
    - *Note: in studies that collect outcomes over multiple time periods (e.g., two year college persistence and then three year college persistence), two year college persistence will be superseded by three year college persistence. In these cases, multiple comparison adjustment may not be required and we encourage you to discuss this with us.*
  - Secondary or exploratory outcomes (as applicable)
  - Definitions of outcomes (e.g., recidivism defined as number of arrests)
  - Data sources, including years of data used
  - Power analysis for primary outcomes, with minimum detectable effects translated into real world terms (e.g., 5 percentage point increase in degree receipt, $350 in earnings)
  - Analysis methods
    - *Description and specification of the proposed empirical model for primary outcomes*
    - *Covariate selection*
    - *Clustering (as applicable)*
  - Key modeling decisions (e.g., years included, functional form, bandwidth selection for RDD, etc.)
  - Key assumptions and proposed tests of assumptions
  - Robustness and sensitivity tests
  - Addressing missing data (as applicable)
  - Addressing data outliers (preferably no trimming)
    - *If trimming, the method for trimming and confirming the blinding of data analysts to which data are treatment vs. counterfactual*

- **Sample Selection**
  - Anticipated sample size and a discussion of inclusion/exclusion criteria
  - Clearly describing the treatment and counterfactual groups (as applicable)