Pre-Analysis Plans
David DuBois, Ph.D.

Research and Professional Development Brown Bag Series
Institute of Health Research and Policy, University of Illinois at Chicago
October 18, 2017

Background

▸ **What is an Analysis Plan?**

  ◦ "a step-by-step plan setting out how a researcher will analyze data which is written in advance of them seeing this data (and ideally before collecting it in cases where the researcher is collecting the data)." (McKenzie, 2012)

  ◦ In principle, could be apply to all types of data, but focus today (and in most discussions to date) will be on quantitive

- Benefits of pre-analysis plans
  - For investigator
    - Can help to refine other aspects of study design
      - Sample size, types of measures, etc.
    - Strengthen funding proposals
      - Reviewers often lament lack of sufficiently detailed plans for data analysis
    - Improved efficiency when time for data analysis comes
      - "Turn-key" plan for statistician or whomever is responsible for data analyses
    - Research assessed as having greater rigor and credibility
      - May translate into publication in higher-impact journals, increased competitiveness for funding of future research, and more favorable assessments of one's work by other researchers (e.g., tenure and promotion reviews)
      - Protection against "second guessing" analytic approach in wake of unexpected findings (e.g., null results) if have potential critics agree on plan beforehand.

"If you don't know where you are going, you'll end up someplace else."
— Yogi Berra

▸ Benefits of pre-analysis plans (cont'd)
  ◦ For science and society
    • Improve transparency in reporting of research
    • More trustworthy/replicable findings
      • Reduce capitalization on chance due to data mining and specification searching (e.g. "p-hacking")
      • Make selective reporting of outcomes more difficult
    • Stakeholders with vested interests -- policy-makers, activists, citizenry, etc. -- can have greater confidence in research findings used to inform their decisions (e.g., help to avoid "policy-based evidence-making")
      • May improve track-record of achieving intended outcomes with research/evidence-informed decisions

UIC

Development of a Pre-Analysis Plan

- **Pre-analysis Plan Checklist**
  - Developed by McKenzie (2012)
- Let's consider the checklist with reference to a "real world" and definitely "imperfect" example:
  - Pre-analysis plan for Youth Relationships Study, an RCT of the Big Brother Big Sisters community-based mentoring program

UIC

Development of a Pre-Analysis Plan (cont'd)

- 1. Description of the sample to be used in the study
  - How will be (or was) sample obtained?
    - Address methods for selection of site(s) if applicable
  - How will representativeness of sample be assessed?
  - Description of randomization procedures if applicable (see CONSORT guidelines for details on what to include)
    - Variables to be included in tests of randomization balance

UIC

Development of a Pre-Analysis Plan (cont'd)

- 2. Key data sources
  - Data collection modes and procedures
  - Especially for survey measures, it may be helpful to develop a measures dictionary that specifies the source of each measure and any adaptations to it (important–these can be forgotten with time!)

Development of a Pre-Analysis Plan (cont'd)

- 3. Hypotheses to be tested
  - ◦ Specify key outcomes of interest
  - ◦ Delineate theory of change/path model that clarifies expected causal chain among outcomes (relates to statistical concept of mediation)
  - ◦ Subgroups to be examined (relates to statistical concept of moderation)
  - ◦ Hypotheses to accompany all tests
    - Useful to distinguish primary and secondary
  - ◦ Emphasis on specificity
    - Not good: A primary outcome will be employment
    - Good: A primary outcome will be employment, as measured by question D21 on the follow-up questionnaire which asks whether the individual currently works for 20 hours or more

UIC

Development of a Pre-Analysis Plan (cont'd)

▶ 4. Specify how variables will be constructed
  ◦ Scoring procedures, including any transformations
  ◦ Handling of missing data (e.g., multiple imputation)
  ◦ Procedures for identifying and handling outliers

Development of a Pre-Analysis Plan (cont'd)

- 5. Specify the treatment effect equation to be estimated (or more generally analytic procedures to be used)
  - Type of procedure (e.g., ANCOVA)
    - Specify fully (e.g., random vs. fixed effect specification if intervention effects estimated over multiple sites)
    - Criteria for aligning type of procedure to be used with observed distribution of outcome measure (e.g., ordinal vs. continuous, Tobit for censored data, zero-inflated Poisson for count variables with large amounts of 0s)
  - Control variables to be included
  - Exact equation
  - Procedures and formula for computing effect sizes
  - Software to be used and specific procedure within it
  - Sensitivity analyses

UIC

Development of a Pre-Analysis Plan (cont'd)

▶ 6. Plan for dealing with multiple outcomes and multiple hypothesis testing
   ◦ Procedures for aggregating different measures into a single index (if applicable)
   ◦ Method to be used for controlling Type I error (e.g., Benjamini-Hochberg)
      • Family-wise level of Type I error (e.g., $p < .05$)
      • What outcomes will be considered part of the same family

UIC

Development of a Pre-Analysis Plan (cont'd)

- 7. Procedures for addressing study attrition
  - Checks that will be done for attrition
    - overall and differential (treatment vs. control) rates
    - representativeness of those with complete data and does degree of representativeness differ for treatment vs. control/comparison group
    - Criteria for what will be considered acceptable
    - Specification of adjustments to be made if problematic attrition is evident (e.g., inverse-probability weighting)

UIC

Development of a Pre-Analysis Plan (cont'd)

- 8. How will the study deal with outcomes with limited variation?
  - Criteria for determining when (if ever) an outcome will be omitted from analysis due to lack of sufficient variation (e.g., everyone in the control group does what the intervention is meant to induce, such as exercise)
    - Including these types of measures can reduce statistical power when included in a family of outcomes
  - Can also plan for omitting variables with levels of non-response above a certain threshold or those whose psychometrics (e.g., coefficient alpha) don't end up reaching an acceptable threshold for the sample

UIC

Development of a Pre-Analysis Plan (cont'd)

- 9. If research will be testing a model, include it
  - Ideally, this would include a path diagram with measures to be used for each construct.
  - Procedures for model testing should be specified as well (see Step 5.; e.g., fit indices/criteria, decision-rules for modifying model specifications such as correlations among error terms, removal of non-significant paths, alternative models against which hypothesized model will be evaluated, etc.)

UIC

Development of a Pre-Analysis Plan (cont'd)

▶ 10. Archive the plan

  ◦ Essential for addressing one of the main purposes of a pre-analysis plan, which is to pre-commit to examining particular measures and outcomes and how this will be done.

  ◦ Several options – one good one is Open Science Framework

  ◦ Note that the plan does not need to be publicly viewable before the study is released – rather, they are time-stamped and released upon request of the researcher

UIC