# Pre-analysis Plan Outline

**Created by Christopher B. Mann**[[1]](#endnote-1)

***Purpose:*** *Outline for pre-registration of analysis plan for political science experiments.*

1. **Research question(s)**

*What empirical question or questions is the experiment designed to answer?*

*Provide definitions of each concept in the question(s) or references to definitions in prior research. This information can be copied from the IRB submission.*

1. **Hypotheses to be tested:** 
   1. Direct hypotheses
   2. Interaction hypotheses (intersection of orthogonal random assignments)
   3. Heterogeneity hypotheses (differences across observational characteristics)

*When stating hypotheses, link each outcome to how it will be measured: (e.g. “outcome will be \_\_\_\_\_, as measured by question \_\_\_ in survey that asks \_\_\_.”)*

*This information can be copied from the IRB submission, but should be organized per the prompts above.*

1. **Description of dataset:** 
   1. Procedures for data collection?

*Describe step-by-step implementation of your experiment with sufficient detail that someone with no other information could replicate the implementation of the experiment.*

* 1. Population Frame
     1. What is the overall intended population?

*Describe the demographics and other characteristics of the intended experimental population.*

* + 1. How is sample collected from overall population? (if applicable)

*Describe the sampling (or recruiting) process used to obtain data from participants.*

* 1. Expected number of observations
  2. Source and type of data in each variable

*List each variable that will be included in the data set, the source of each variable, the type of each variable (survey, administrative, etc), and how different sources will be matched together.*

1. **Specify how variables will be processed:** 
   1. Describe turning raw data into data for analysis

*Detail how raw data (#3.4 above) will be altered into variables to be used in the analysis.*

* + 1. Special attention to:
       1. Difference-in-difference outcomes
       2. Indices (outcome or covariates)
       3. Transformation of variables (e.g. bins or dummies from continuous data, log, quadratic, etc.)
       4. Imputation of missing data
  1. What steps will be taken to keep the processed data confidential?
     1. What potential personally identifying information is in collected data
     2. When and how will identifying information be removed from dataset
     3. How will data be secured (before and after PII is removed)

*Note: Best practice is providing actual code to convert raw data to analysis variables.*

1. **Randomization:**
   1. **Process**
      1. Method of random assignment

*Describe step-by-step implementation of your experiment with sufficient detail that someone with no other information could replicate the implementation of the experiment. Be clear about when random assignment will occur: pre-treatment based on a list or at time of treatment.*

* + 1. Number of randomly assigned conditions
       1. Number of treatment conditions

*State whether treatment conditions will be pooled (e.g., any treatment or theoretically informed combinations of treatments such as row vs. columns in 2x2 design) or always analyzed separately*

* + - 1. Control, placebo, or both?
    1. Blocking of randomization

*List any covariates used to create blocks in the randomization*

* + - 1. Probability of assignment within blocks

*List probability of assignment for each block or state that probability of assignment is the same for all blocks.*

* + 1. Level of random assignment in relation to data structure
       1. Clustering for household, jurisdiction, or other multi-record unit?
    2. Re-randomization
       1. Criteria to reject/accept random assignment (if any)
       2. Procedure used for re-randomization (e.g. program or code)
  1. **Balance**
     1. What statistical procedure and summary statistics used for balance checks
     2. What variables used in pre-treatment checks of balance
     3. What variables used in post-outcome checks of balance after attrition

1. **Estimating Treatment Effects** 
   1. **Type of treatment effect estimation:**
      1. Intent to treat

*Note: I believe you should always calculate the ITT, but note if this is not the quantity of interest and therefore relegating it to supplemental info*

* + 1. Conditional average complier effect (a.k.a. treatment on the treated)
       1. Specify method of calculating (e.g. instrumental variables vs. dividing by contact rate)
  1. **Specify each treatment effect equation to be estimated:**

*Note: The exact equation to be estimated should be written out.*

* + 1. What type of analysis will be done? (e.g., t-test, regression, etc.)
  1. **Specify how multiple treatments will be compared?** (if applicable)
     1. What type of analysis will be done? (e.g., f-test of regression coefficients)
  2. **How will standard errors be calculated?**

*Note: Standard error calculation must be consistent with level of randomization*

* 1. **What covariates included in each equation?**
     + 1. Block variables
       2. Other covariates
  2. **Heterogeneous Treatment Effect estimation**
     1. Is heterogeneity based on a block variable in randomization? (i.e., interaction hypotheses vs heterogeneity hypotheses)
     2. What type of analysis will be done? (e.g. interaction terms)

1. **Dealing with multiple outcomes and multiple hypothesis testing?** 
   1. Describe which outcomes are measures of the same concept (i.e. tests of the same hypothesis) for making inferences?
   2. Which (if any) outcome measures will be grouped in an index for hypothesis testing?
2. **Dealing with problems in post-outcome data**
   1. **Procedures for addressing attrition:** 
      1. What checks will be done for attrition? (e.g. post-outcome balance check)
      2. What adjustments will be made for selective attrition?
   2. **Too little variation in outcome measure** 
      1. Decision rule / threshold on dropping outcome measure from analysis
   3. **Missing data / item non-response in an outcome measure** 
      1. Decision rule / threshold on dropping outcome measure from analysis
   4. **Missing data / item non-response in a covariate**
      1. Decision rule / threshold for dropping covariates from analysis equations
3. **Pre-register analysis plan** 
   1. File plan with a binding registry for transparency/credibility with scholarly community

*Note: For class, turning in the assignment is sufficient ‘pre-registration’. If you plan to use this data beyond class for a thesis, conference paper, etc., speak to me about pre-registration at OSF.*

**Endnotes**

1. Outline based on my experience and the following sources:

   Mackenzie, David. 2012. “A Pre-Analysis Plan Checklist.” Impact Evaluations @ the World Bank. October 28. <https://blogs.worldbank.org/impactevaluations/a-pre-analysis-plan-checklist>

   Ganiman, Alejandro. 2014. “Pre-Analysis Plan Template.” <http://cega.berkeley.edu/assets/cega_events/92/Pre-Analysis_Plan_Template_Alejandro_Ganimian.pdf> [↑](#endnote-ref-1)