



# A Guide to Pre-Analysis Plans

Erica Chuang and Stephanie Wykstra, Research and Knowledge Management

Innovations for Poverty Action, 2015

## Introduction

As a part of a movement towards greater research transparency, J-PAL and IPA have adopted pre-registration as a requirement for all projects. While pre-registration on the AEA registry does not require a pre-analysis plan, a number of researchers decide to undertake pre-analysis plans (PAPs).

In some cases, IPA research staff have also been asked by PIs to contribute to writing pre-analysis plans. The Research and Knowledge Management team at IPA has created this guide to provide guidance in writing a pre-analysis plan for IPA staff (as well as interested researchers).

The guide consists in an **introduction to pre-analysis plans**, a description of the **items that should be included**, and finally, a **checklist** to use while writing a plan.

Please contact [researchsupport@poverty-action.org](mailto:researchsupport@poverty-action.org) with any questions or comments.

## Pre-Analysis Plans: What, When and Why?

### What?

A pre-analysis plan (PAP), sometimes called a research or study protocol, is a document outlining the technical details of a study, written before the study is conducted. Included in a PAP is information such as the type of study to be conducted, sources of data, how the variables will be constructed, model specifications, problems that may arise over the course of the study and how those problems will be addressed.

### Why?

The primary importance and objective of PAPs/study protocols are to raise the credibility and reliability of research results. PAPs allow for a higher level of confidence, particularly where researchers are seeking to confirm a hypothesis rather than simply explore potential causal relations. By “tying one’s

hands” through a PAP, it is possible for both the researchers and others to have a higher level of confidence that reported results do not reflect statistical noise hand-picked from the data.

Research protocols have often been used in the natural and physical sciences when developing new technologies, drugs, or treatments. Medical research journals often require pre-registration of clinical trials, including a protocol describing primary and secondary outcomes, as a prerequisite for publication ([ICJME, Clinical Trials Registration](#)). In recent years, advocates within the social sciences also have promoted the adoption of pre-analysis plans (Miguel et al. 2014, BITSS forum 2013, Casey et al. 2012).

## When?

A PAP should be written and registered **before** the intervention begins, so as to ensure and publicly display the fact that a hypothesis is made prior to seeing any data. There may be exceptions to this rule, but our guide is geared towards studies for which PAPs may be written before data collection.<sup>1</sup>

## What should be included in a Pre-Analysis Plan:

Below we suggest and discuss elements to include in a PAP. The guidelines were written with social science RCTs in mind, but they may be applied to other study methods as well.

### Basic information/introduction:

The PAP should make it clear who is working on the study, and the objective for undertaking it.

- **Name/title** of the project
- **Authorship:** all researchers involved and their affiliations.
- **External partner institutions** (if applicable)
- **Conflicts of Interest:** list any conflicts of interest for all study authors
- **Brief summary of project and objective/rationale** for conducting this study.

### Study Design:

#### Hypotheses and how treatment effects will be measured:

- Include all hypotheses which will be tested.
- What are the main variables of interest, and how will the treatment effects be measured?
- Which outcomes are primary to the analysis, which are secondary, and why?

---

<sup>1</sup> In some cases, it may not be possible to lay out a pre-analysis plan prior to data collection. Glennerster and Takarvarasha suggest that if a PAP cannot be written before data has been collected, then at the very least it should be done before any analysis has been conducted and any data has been manipulated. In that case, we’d recommend clearly indicating when the plan has been written with respect to the study timeline (i.e., after midline). In addition, there may be cases in which it is appropriate to write PAPs after distinct waves of treatment, as is the case with the Generasi study included in the J-PAL Hypothesis Registry (see Generasi 2014).

**Preliminary studies (if applicable):** If this project is part of a series of studies, provide a brief description of the prior studies and how they were relevant to generating the hypotheses.

**Basic methodology:**

- What is the methodology of the study (RCT, natural experiment, regression discontinuity design, etc)?
- Is the study blind, double-blind or unblinded?

**Describe the study design and procedures including:**

- The geographic region
- Research population
- Sampling frame
- Inclusion/exclusion criteria (with clear justification)
- Unit of analysis and definition of cluster (if applicable)
- Attrition criteria as applied to individual participants
- Early termination criteria for the study (e.g. the treatment works so well that the project is halted and the treatment is administered to all study participants)
- Expected timeline of the study, including a detailed description of when the intervention and data collection will take place

**Describe the intervention in detail:** Describe the intervention in detail. If the intervention involves a technology or medical treatment, provide technical details for how the technology/treatment will work. Provide a description of how the intervention will differ between the control and treatment groups and/or between subgroups (if applicable). If there are several treatment arms, detail how each arm will differ from each other as well as the control group.<sup>2</sup>

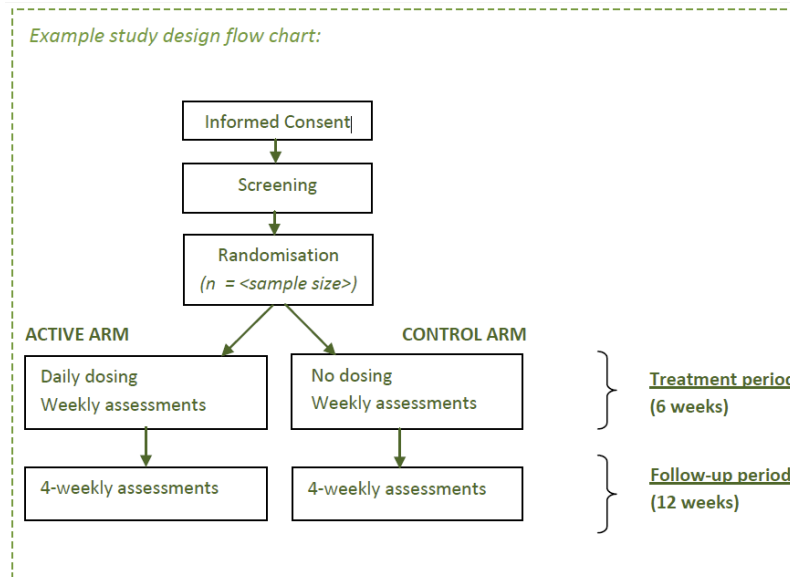
**Include a flow chart (optional):**

It may be helpful to include a flow chart/diagram illustrating the approximate timeline and/or structure of the project, though not necessary. Below is an example of such a flow chart from the Murdoch Children's Research Institute.<sup>3</sup>

---

<sup>2</sup> If there are several waves of treatment in the study and only one PAP will be written, indicate the approximate intervals in which each wave will occur, any changes made to implementation in each wave, etc.

<sup>3</sup> Note: initial number of participants assigned to each treatment arm should also be specified in a flow chart, if known. For source of example diagram, see Murdoch Children's Research Institute 2009.



### Data collection and sources of data:

What will the sources of data be? How will the data be collected (e.g., in-person interviews)? Provide rationale for using certain methods of data collection over others. If there is data used in the study that is not collected by your team (e.g., third-party administrative data), describe the source of the data.

### Randomization:

If the study is a randomized controlled trial, define the randomization procedure in advance.

- How will it be done (In excel? Physical lottery? Through a statistical program such as STATA?). If there is a process with pieces of paper randomly drawn, how will the process ensure that it is not possible to interfere with randomization?
- What is the unit of randomization (e.g. individual, cluster)? Provide justification for this decision.
- What stratification variables are used (if any)?

### Blinding (if applicable):

Please state who will be blinded (participants, survey staff, data analysts, etc.), and procedures in place to ensure only those who are authorized to know of the treatment status of the participants have access to that information.

### Power calculations

Power calculations justifying sample size should be included. Justification for the effect size included in the calculation should also be noted. **Note:** IPA recommends [Optimal Design](#), a software tool that is designed for running power calculations, as well as Stata. Here we link to J-PAL's instructional guide for how to use Optimal Design.

## Specifying analytical decisions in advance<sup>4</sup>:

*For further detailed guidance on this sub-topic, see McKenzie 2012.*

- If variables will be constructed (e.g. creating index variables), how will they be constructed?
- What are the primary model specifications? Include this in the PAP.
- If multiple hypothesis testing will be done, how will this be accounted for in the analysis?

## Accounting for Issues

Note other relevant issues that may arise during implementation, and the plan for dealing with them. For example, how will non-compliance of participants and/or contamination be monitored and accounted for?

## Registering the trial

It is suggested that prior to beginning the project, register the study in a public trial registry. These sites ensure that the study and plan was registered on a particular date, prior to data collection.

J-PAL and IPA require that studies are registered in the AEA registry (see link below). There are also other registries, noted here for reference.

## Resources

- The [AEA registry](#) includes the option to upload PAPs. Search under “advanced options” for studies which include PAPs.
- Note that the older The [J-PAL Hypothesis Registry](#), superseded by the AEA registry, has a collection of pre-analysis plans of RCTs for both economics and political science from 2009 to 2013.
- [EGAP](#) is a registry for political science studies, some of which include pre-analysis plans.
- [3ie](#) also has a database of ongoing international development impact evaluations.
- The [Open Science Framework](#) (OSF) also invites pre-registered studies and PAPs.
- Many clinical trials in the US must be registered with [Clinicaltrials.gov](#).
- [PubMed](#) is searchable by “protocols” for many different kinds of studies.
- [SPIRIT](#) is a protocol checklist often used in the natural sciences.

---

<sup>4</sup> See David McKenzie’s “Pre-analysis Plan Checklist” for more detail on questions in this section. (McKenzie 2012.)

## Pre-analysis Plan Checklist

Below is a checklist for all of the sections and elements to be included in a PAP. Optional information is highlighted in **blue**, and information to include if applicable is highlighted in **green**.

### ☐ **General information**

- ☐ Title of the project
- ☐ Researchers involved
  - ☐ Name
  - ☐ Title
  - ☐ Degree (optional)
  - ☐ Department
  - ☐ Institution
- ☐ External partner institutions (if applicable)
- ☐ Project staff (optional)
- ☐ Conflicts of interest (if applicable)

### ☐ **Introduction**

- ☐ Project Summary/Abstract
- ☐ Aims, rationale and background

### ☐ **Study Design**

- ☐ Hypotheses
- ☐ Treatment effects and measurement
  - ☐ Main variables of interest, precise definitions
  - ☐ How outcomes will be measured
  - ☐ Distinction between primary and secondary outcomes
- ☐ Preliminary studies (if applicable)
- ☐ Details of study:
  - ☐ Geographic region
  - ☐ Research population
    - ☐ Demographic information on target populations/subgroups
    - ☐ Clear rationale for inclusion or exclusion of certain populations
    - ☐ Procedures for recruitment and consent
    - ☐ Potential benefits and risks to subjects
  - ☐ Sampling frame
  - ☐ Inclusion/exclusion criteria
  - ☐ Withdrawal criteria
  - ☐ Early termination criteria
  - ☐ Expected timeline
  - ☐ Treatment waves (if applicable)
- ☐ Intervention
  - ☐ Technical components of the intervention (e.g. drug or technology).
  - ☐ Differences between the treatment and control arms

- ☐ Differences between each distinct treatment arm (if applicable)
  - ☐ Definition of cluster and differences between cluster and unit of analysis (if applicable)
  - ☐ Flow chart diagramming treatment arms, sample size and timeline (optional)
- ☐ Data collection methods and procedures
  - ☐ Description of data collection method
  - ☐ Description of any other data sources used (e.g. admin data) and source of data
- ☐ Randomization procedure
  - ☐ Detailed description of the mechanism for randomizing, including how the process will be safeguarded from tampering
  - ☐ Individual vs. clustered randomization?
  - ☐ Stratification variables?
- ☐ Blinding
  - ☐ Describe blinding (who and how) (if applicable)
- ☐ Power calculations
  - ☐ Justification of effect size used in power calculations
- ☐ **Pre-specifying analytical decisions:**
  - ☐ Type of model and justification for its use (if applicable)
  - ☐ Variables to be constructed – describe details of construction (if applicable)
  - ☐ Accounting for multiple hypothesis testing (if applicable)
  - ☐ Any indices used, mathematical formulae, explanation and rationale (if applicable)
- ☐ **Expected issues**
  - ☐ Details of procedures in place to address issues e.g., how will non-compliance be monitored and accounted for?
- ☐ **Conclusion**

## References:

- Berkeley Initiative for Transparency in the Social Sciences (BITSS), 20 Mar. 2013. "An Open Discussion on Promoting Transparency in Social Science Research." <<http://bitss.org/2013/03/20/an-open-discussion-on-promoting-transparency-in-social-science-research/>>.
- Casey, K., R. Glennerster, and E. Miguel. "Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan." *The Quarterly Journal of Economics* 127.4 (2012): 1755-812.
- Chan A-W, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, Hróbjartsson A, Mann H, Dickersin K, Berlin J, Doré C, Parulekar W, Summerskill W, Groves T, Schulz K, Sox H, Rockhold FW, Rennie D, Moher D. *SPIRIT 2013 Statement: Defining standard protocol items for clinical trials*. *Ann Intern Med*, 2013;158:200-207.
- Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin J, Dickersin K, Hróbjartsson A, Schulz KF, Parulekar WR, Krleža-Jerić K, Laupacis A, Moher D. *SPIRIT 2013 Explanation and Elaboration: Guidance for protocols of clinical trials*. *BMJ*, 2013;346:e7586.
- CONSORT Group "CONSORT Statement 2010." <<http://www.consort-statement.org/>>.
- Gertler, Paul, and Sebastian Martinez. *Impact Evaluation in Practice*. Washington, D.C.: World Bank, 2011. Print.
- Glennerster, Rachel, and Kudzai Takavarasha. *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ: Princeton UP, 2013. Print.
- McKenzie, David. "A Pre-analysis Plan Checklist." The World Bank, 2012.  
<<http://blogs.worldbank.org/impactevaluations/a-pre-analysis-plan-checklist>>.
- Miguel, E., C. Camerer, K. Casey, J. Cohen, K. M. Esterling, A. Gerber, R. Glennerster, D. P. Green, M. Humphreys, G. Imbens, D. Laitin, T. Madon, L. Nelson, B. A. Nosek, M. Petersen, R. Sedlmayr, J. P. Simmons, U. Simonsohn, and M. Van Der Laan. "Promoting Transparency in Social Science Research." *Science* 343.6166 (2014): 30-31.
- Murdoch Childrens Research Institute, 2013. "Research Resources."  
<<http://www.mcric.edu.au/research/core-facilities/clinical-research-development-office/other-research-resources/>>.
- Olken, Benjamin, Junko Onishi and Susan Wong. 2014. "Project Generasi: Conditional community block grants in Indonesia." *AEA RCT Registry*. May 15.  
<<https://www.socialscienceregistry.org/trials/332/history/1759>>.



University of Oklahoma Health Sciences Center. "Research Protocol Outline."

<[https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB4QFjAA&url=https%3A%2F%2Fwww.ouhsc.edu%2Ffirb%2Ffirb\\_Forms%2Ffirb%2FResearch\\_Protocol\\_Outline\\_2282007.doc&ei=HqrKVNqlEMf8ggT0s4OgAQ&usg=AFQjCNGsFs0fSDI26ILYeUFNH7kfJCidhw](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB4QFjAA&url=https%3A%2F%2Fwww.ouhsc.edu%2Ffirb%2Ffirb_Forms%2Ffirb%2FResearch_Protocol_Outline_2282007.doc&ei=HqrKVNqlEMf8ggT0s4OgAQ&usg=AFQjCNGsFs0fSDI26ILYeUFNH7kfJCidhw)>.

World Health Organization. "Recommended Format for a Research Protocol."

<[http://www.who.int/rpc/research\\_ethics/format\\_rp/en/](http://www.who.int/rpc/research_ethics/format_rp/en/)>.