



Berkeley Initiative for  
Transparency in the Social Sciences



---

## Reproducibility & Transparency Workshop

Maanzani Lodge, Athi River, Kenya

March 7–8, 2016



## Background

The social science research community is increasingly emphasizing research transparency as an essential component of high quality research. Research transparency includes study pre-registration, publishing pre-analysis plans (PAPs), following standard reporting guidelines, publicly sharing data and underlying code, and other practices that enable reproducible research.

[Innovations for Poverty Action \(IPA\)](#) and the [Center for Effective Global Action \(CEGA\)](#) are research hubs that use randomized control trials (RCTs) because they provide the highest quality and most reliable answers to what works and what does not. The [Berkeley Initiative for Transparency in the Social Sciences \(BITSS\)](#)<sup>1</sup> and the [Center for Open Science \(COS\)](#) are key actors in the research transparency movement. COS provides free and open services to increase transparency and reproducibility of research, while BITSS develops and delivers education and training on the range of research transparency tools and methods. Both BITSS and COS are working to realign incentives and practices with scientific values of openness, integrity, and transparency across the social sciences.

## Problem and Motivation

One of the main goals in both IPA and CEGA is to support high quality, transparent, reproducible research. The research support staff at both organizations is typically tasked with preparing data and code for publication. However, key steps for preparing data and code take place as early as the study design phase as well as in the data collection process.

As such, many activities required during data production and documentation are informed by these early stages, including: 1) organizing file structure, 2) creating understandable variable labels and value codes, as well as connecting variables to survey instruments through consistent labels and codebook creation, 3) version control of code and data, 4) creating and maintaining documentation files about the survey and data, as well as data cleaning steps.

Therefore, both IPA and CEGA realize that training on research transparency tools and methods early in the research process, and applying those practices throughout, will enable more effective production of transparent, reproducible research.

---

<sup>1</sup> BITSS is an initiative that is based within the Center for Effective Global Action (CEGA) at UC Berkeley

## Objectives

We have several goals for the research transparency workshop:

1. Familiarize participants with the importance of reproducible research and research transparency (both for their own work as well as from a public good perspective).
2. Present concrete guidance and tools e.g. recommendations from IPA's best practice manual for managing data/code, BITSS' reproducible research handbook, Github, Open Science Framework and Dataverse.
3. Provide an excellent setting with support for participants to practice new skills / tools, and to improve their own data/code from their projects.

## Audience

We propose to offer a pilot workshop for research staff including Research Analysts, Coordinators, and Managers working with IPA, J-PAL, and CEGA, as well as others based in developing world research groups and universities. By training staff and researchers early in the research lifecycle, we aim to promote reproducible research with concrete guidance about how to make materials understandable for publication.

Given this is a pilot, IPA and BITSS will send targeted invitations for up to 30 participants who meet the following criteria:

- Master's level (either completed, or ongoing) or higher in economics, public policy, statistics, or other relevant degree
- Training and experience with statistics
- Experience with and currently using STATA and/or R for their own research

## Agenda

This will be a two-day workshop,  
with the following general agenda.

### **March 7, 2016**

- |                   |                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 09:00am - 09:15am | Welcome and Introductions                                                                                                                                                                                                                                                                                                                                                                                                                               |
| 09:15am - 12:00pm | Research Transparency and Reproducible Research - What, Why, and How<br>Presentations covering high-level topics such as registration, data publication and replication to familiarize participants with the motivations for transparency and reproducible research, as well as current topics in these areas;<br>Separate into three sessions: on registration and pre-analysis plans, data-sharing and replication, and software for reproducibility. |
| 12:00pm - 01:00pm | Lunch                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
| 01:00pm - 03:00pm | Concrete Guidance on Making Research Data and Code Reproducible<br>Overview of concrete principles such as file structure, consistent variable labeling, code commenting, version control, readme files, and codebooks;<br>Hands-on practice exercises with data and code provided by organizers.                                                                                                                                                       |
| 03:30pm - 05:00pm | Statistical Programming<br>Basic coding in Stata/R, for participants with relatively little experience.<br><br>Pre-Analysis Plans<br>Tutorial in how-to of writing a pre-analysis plan.                                                                                                                                                                                                                                                                 |
| 05:30pm - 07:30pm | Dinner<br>For all workshop participants and staff.                                                                                                                                                                                                                                                                                                                                                                                                      |

## **March 8, 2016**

09:00am - 10:15am	<b>Git for Version Control</b> Hands-on practice with Git to familiarize participants with tool for version control on their own computers. Brief introduction to Github as well.
10:30am - 11:45am	<b>Open Science Framework</b> For collaborative workflow management.
11:45am - 12:45pm	<b>Reproducible Research in the Field</b> Visiting speakers participate on a panel, where have they seen code/data management practices implemented well vs. poorly, and what were the implications for research quality and reproducibility?
12:45 - 01:45pm	<b>Lunch</b>
01:45pm - 03:45pm	<b>Hands-on Work with Data/Code</b> For those with projects currently running: checklist of items to run through, following Day 1 principles for managing data/code. TAs available to answer questions; For those without current data/code from projects: following checklist using example data/code provided by organizers.
03:45 - 04:00pm	<b>Certificates and Closing</b>