

**Replication kit for the paper:**  
**“The Challenges of Universal Health Insurance in Developing Countries:  
Experimental Evidence from Indonesia’s National Health Insurance”**

Abhijit Banerjee, MIT  
Amy Finkelstein, MIT  
Rema Hanna, Harvard University  
Benjamin A. Olken, MIT  
Arianna Ornaghi, University of Warwick  
Sudarno Sumarto, TNP2K and SMERU

This is the documentation for the replication kit for the paper “The Challenges of Universal Health Insurance in Developing Countries: Evidence from a Large Scale Randomized Experiment in Indonesia.” It contains a data availability statement, detailing the source and accessibility of the data used in the paper, as well as replication instructions.

## **1. Data Availability and Provenance Statement**

### **Statement about Rights**

The authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

### **Summary of Availability**

Survey data with personally identifiable information and the administrative data **cannot be made** publicly available.

### **Details on each Data Source**

The projects combines five main data sources, outlined below.

1. Survey data (Banerjee et al. 2021)

For this paper we collected a short baseline survey in conjunction with an independent and established survey firm (SurveyMeter). The survey instruments used in the two study sites are provided in Indonesian and English in the subdirectory `survey_materials/`. These can be used as a dictionary for the variables.

We are able to provide partial access to the data collected. In particular, we are not able to fully share the following sections. First, we cannot provide access to all variables in section AR, because they include personally identifiable information. We instead provide access to a version of the data that is stripped from PII. Second, for the same reason, we cannot provide access to section LK (which includes detailed location information). Third, we

cannot provide access to section VC, because the data agreement with BPJS prevents us from disclosing the treatment status of different households.

All other sections are provided in `data/raw/baseline_survey_bandung_public/` and `data/raw/baseline_survey_medan_public/`, in dta format.

Note that we also collected data during the implementation of the project. In particular, the analysis uses data on what happened when households visited the local BPJS office. Because the data contains personally identifiable information, they are confidential and cannot be publicly shared.

## 2. Administrative data from BPJS (BPJS, 2021)

The data measures enrollment, payments, and claims using administrative data from Badan Penyelenggara Jaminan Sosial – Kesehatan (Social Security Administration for Health, or BPJS). The datasets used in the analysis were created by BPJS by matching the national ID numbers (Nomor Induk Kependudukan, or NIK) of the study participants to BPJS's enrollment, payments, and claims databases. The data are confidential, but may be obtained with Data Use Agreements with BPJS. It can take some months to negotiate data use agreements and gain access to the data. Researchers interested in access to the data may contact:

Citra Jaya

Deputy Assistant for National Health Security Research

BPJS Kesehatan

JL Letjen Suprpto Cempaka Putih

PO BOX 1391 JKT-INDONESIA 10510

+62 21-4212938

[citra.jaya@bpjs-kesehatan.go.id](mailto:citra.jaya@bpjs-kesehatan.go.id)

## 3. Indonesian National Household Survey March 2015 (BPS, 2015)

The data from the March 2015 Indonesian National Household Survey (BPS, 2015) are proprietary, but may be purchased from BPS at the following [link](#). To purchase the March 2015 data, click on the “Penjualan” tab. This connects to a [login page](#), where it is possible to create an account. After logging in, the relevant data can be found and purchased by clicking on the “microdata” section and following the instructions provided. Note that both the KOR data and the Expenditure and Consumption modules need to be purchased to replicate the analysis.

## 4. ICD10 codes (WHO, 2008)

The ICD-10 Version:2008 codes were obtained at the following [link](#), and are provided in `data/raw/icd10_confidential/` in dta format.

5. Chronic disease categorization (Agency for Healthcare Research and Quality, 2018)

Data on whether a condition is chronic or not is from the Chronic Condition Indicator for the International Classification of Diseases from the Healthcare Cost and Utilization Project (HCUP, 2018). We use version 2018/1. The data can be downloaded at the following [link](#), and are provide in `data/raw/icd10_confidential/` in dta format.

## Dataset List

See `README_tables.xlsx`, sheet “dataset\_list.”

## 2. Replication Instructions

### Software Requirements

- Stata (code was last run with version 15)
  - `carryforward` (as of 2017-09-11)
  - `estadd` (as of 2017-09-11)
  - `estout` (as of 2019-08-19)
  - `eststo` (as of 2016-04-25)
  - `esttab` (as of 2017-09-11)
  - `mdesc` (as of 2015-07-31)
  - `reclink` (as of 2017-04-27)
  - `estout` (as of 2018-05-12)

Note that all ado files needed to run the analysis are provided as part of the replication package (in the folder `code/ado/`). Calling the do file `code/01_setup` ensures that Stata calls the programs from the correct directory.

### Memory and Runtime Requirements

Approximate time needed to reproduce the analyses on a standard desktop machine: less than ten minutes.

### *Details*

The code was last run on an Intel(R) Core(TM) i5-5200U CPU @ 2.20GHz 2.19GHz with 8.00 GB of installed memory (RAM) with Windows 10 Enterprise.

## Replication Folder Structure

The replication folder is organized in the following subdirectories:

- The subdirectory `code/` includes all code necessary to replicate the analysis.
- The subdirectory `data/` contains all the data. The `data/raw/` subdirectory contains the raw data files, the `data/working/` subdirectory is an empty subdirectory to save all intermediate files created during the data cleaning, and the `data/codeddata/` subdirectory is an empty subdirectory to save the final datasets used in the analysis.
- The subdirectory `output/` is an empty subdirectory to save all figures (`output/figures/`) and tables (`output/tables/`).
- The subdirectory `survey_materials/` contains the listing and baseline surveys.

## Description of Code

All code necessary to replicate the analysis is provided in the subdirectory `code/`. The subdirectory contains three Stata do files (named: `code/01_setup.do`, `code/02_master_cleaning.do`, `code/03_master_analysis.do`) and three subdirectories (`code/ado/`, `code/analysis/`, `code/cleaning/`).

- The do file `code/01_setup.do` correctly sets the system and working directories.
- The subdirectory `code/ado/` contains the ado file necessary to run the analysis. The do file `code/01_setup.do` sets the ADO directories appropriately.
- The subdirectory `code/cleaning` contains all do files necessary to clean the data. The do file `code/02_master_cleaning.do` will run them all, in the correct order. The do file `code/02_master_cleaning.do` also briefly describes what each do file does.
- The subdirectory `code/analysis` contains all do files necessary to clean the data. The do file `code/03_master_analysis.do` will run them all. Each do file called from `code/03_master_analysis.do` identifies the type of output that it creates (e.g. `code/analysis/tables.do` creates the tables). Output files are called appropriate names (`table5.csv`, `figure6.pdf`) and should be easy to correlate with the manuscript.

## Instructions to Replicators

The code provided constructs all the cleaned analysis datasets from raw datasets. Due to data restrictions, we cannot provide all the raw datasets used, nor the derivative cleaned analysis datasets.

For replication that does not involve restricted data: We provide the raw data, and the do files to create the intermediate data from the raw data.

Steps for replication with the non-restricted data only:

- Edit `code/01_setup.do` to adjust the default path and set up the correct system directories.

- Run the do file `code/cleaning/cleans_baseline_survey_bandung.do` and the do file `code/cleaning/cleans_baseline_survey_medan.do` to clean the survey data. These are the only data that can be made publicly available.

For replication that involves restricted data: We provide the do files that produce the analysis data from the raw data, and the do files that use the analysis data to produce the final output. Note, in order for these do files to run, it is necessary to first obtain access to the restricted raw data sets.

Steps for replication with the restricted data:

- Mount the data with PII that are stored in an encrypted volume to drive A: /.
- Edit `code/01_setup.do` to adjust the default path and set up the correct system directories.
- Run the do files (in this order) `code/01_setup.do`, `code/02_master_cleaning.do`, `code/03_master_analysis.do` to run all steps in sequence.

### ***Details***

Note that the detailed steps only refer to replication with the restricted data.

- `code/01_setup.do`: will adjust the default path and set up the correct system directories.
- `code/02_master_cleans.do`: will call all the necessary files to clean the data in the correct order.
  - This do file was last run in April 2021.
  - All do files called by the master cleaning do file are in the subdirectory `code/cleaning/`. All programs can be run individually; but order is important.
  - Running the do file should take about 5 minutes.
- `code/03_master_analysis.do`: will call all the necessary files to analyze the data.
  - This do file was last run in April 2021.
  - All do files called by the master cleaning do file are in the subdirectory `code/analysis/` subdirectory. All programs can be run individually; order is not important.
  - Running the do file should take about 3 minutes.

### **List of Tables and Programs**

The provided code reproduces all statistics, figures, and tables in the paper (upon access to the restricted data).

See `README_tables.xlsx`, sheet “output\_list.”

### 3. References

**Agency for Healthcare Research and Quality**, “Healthcare Cost and Utilization Project (HCUP) Chronic Condition Indicator,” 2018. [www.hcup-us.ahrq.gov/toolssoftware/chronic/chronic.jsp](http://www.hcup-us.ahrq.gov/toolssoftware/chronic/chronic.jsp).

**Badan Penyelenggara Jaminan Sosial Kesehatan (BPJS)**, “Administrative Data on Membership, Payments, and Claims [database],” 2021. Jakarta, Indonesia, last accessed 2018-08-16.

**Badan Pusat Statistik (BPS)**, “Survei Sosial Ekonomi Nasional (SUSENAS),” March 2015.

**Banerjee, Abhijit, Amy Finkelstein, Rema Hanna, Benjamin A. Olken, Arianna Ornaghi, and Sudarno Sumarto**, “Data and code for: The Challenges of Universal Health Insurance in Developing Countries: Experimental Evidence from Indonesia’s National Health Insurance,” 2021. American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor], <http://doi.org/10.3886/E138701V1>.

**World Health Organization (WHO)**, “International Statistical Classification of Diseases and Related Health Problems 10th Revision Version: 2008,” 2008. <https://icd.who.int/browse10/2008/en#/>.