Documentation file for Do-files and Data

"Can transparency and accountability programs improve health? Experimental evidence from Indonesia and Tanzania"

Jean Arkedis[1]
Jessica Creighton[2]
Akshay Dixit[3]
Archon Fung[4]
Stephen Kosack[5]
Dan Levy[6]
Courtney Tolmie[7]

Folder organization

Materials on two experiments, organized by country.

T4D

- T4D pre analysis plan
- analysis.do
- Indonesia
  - ➔ Analysis
  - ➔ Data
  - ➔ Do_files
  - ➔ Questionnaires
- Tanzania
  - ➔ Analysis
  - ➔ Data
  - ➔ Do_files
  - ➔ Questionnaires

Note: The folders labelled "Analysis" are empty and meant to store output produced by the .do files.

[1] Results for Development, email: jarkedis@r4d.org
[2] Harvard University, email: jessica_creighton@hks.harvard.edu
[3] Harvard University, email: akshay_dixit@g.harvard.edu
[4] Harvard University, email: archon_fung@hks.harvard.edu
[5] University of Washington, email: skosack@uw.edu
[6] Harvard University, email: dan_levy@hks.harvard.edu
[7] Results for Development, email: ctolmie@r4d.org

Questionnaires

Two questionnaires in English for each of Indonesia and Tanzania, in portable document format. These were used for the post-intervention (endline) data collection.

- *facility_survey_indonesia*: Questionnaire administered to the staff of 200 sample *puskesmas* in Indonesia.
- *household_survey_indonesia*: Questionnaire administered to 6,001 recent mothers across 200 sample villages in Indonesia.
- *facility_survey_tanzania*: Questionnaire administered to the staff of 153 sample dispensaries in Tanzania.
- *household_survey_tanzania*: Questionnaire administered to 6,008 recent mothers across 200 sample villages in Tanzania.

Data

All datasets listed below were deidentified. Personally identifying information, and combinations thereof, were removed to protect the identities of individuals. All datasets are in comma delimited (.csv) as well as Stata dataset (.dta) formats, with the exception of *pvalues_intermediate_outcomes,* which is in Excel (.xlsx) format.

Indonesia

- *facility_data*: Facility-level dataset containing information on post-intervention characteristics of the sample *puskesmas* as described in *facility_survey_indonesia* and the treatment assignment variable.
- *household_data*: Household-level dataset containing information from the post-intervention survey following *household_survey_indonesia* and the treatment assignment variable.
- *household_data_with_main_outcomes*: Household-level dataset containing all the information in *household_data,* plus main (primary and secondary) outcomes cleaned and prepared for regression analyses.
- *Household_Merged_BL_IN*: Household-level dataset containing information from the pre-intervention survey of recent mothers. Note that the study followed a repeated cross-section design, and independent samples of households were selected from each sample village in the pre- and post-intervention surveys.
- *Facility_BL_IN*: Facility-level dataset containing information on pre-intervention characteristics of the sample *puskesmas*. Variable labels are in Bahasa.
- *randomization_strata*: Treatment status and randomization strata for sample villages.
- *pvalues_intermediate_outcomes*: List of p-values from the analysis of all intermediate outcomes, for the purpose of adjustment for multiple hypotheses testing.

Tanzania

- *T4D_end_fc_clean:* Facility-level dataset containing information on post-intervention characteristics of the sample dispensaries as described in *facility_survey_tanzania* and the treatment assignment variable.
- *T4D_end_hh_clean:* Household-level dataset containing information from the post-intervention survey following *household_survey_tanzania* and the treatment assignment variable.
- *T4D_end_hh_with_main_outcomes:* Household-level dataset containing all the information in *T4D_end_hh_clean,* plus main (primary and secondary) outcomes cleaned and prepared for regression analyses.
- *bl_FCMain:* Facility-level dataset containing information on pre-intervention characteristics of the sample dispensaries.
- *bl_HHData:* Household-level dataset containing information from the pre-intervention survey of recent mothers. Note that the study followed a repeated cross-section design, and independent samples of households were selected from each sample village in the pre- and post-intervention surveys.
- *bl_HHBirth:* Birth-level dataset containing information from the pre-intervention survey of recent mothers. In contrast to *bl_HHData* in which rows corresponds to households, in this dataset rows correspond to births in sampled households.
- *sample_villages:* List of sample village IDs and their treatment status.
- *randomization_strata:* Randomization strata for each sample village.
- *sample_facility:* Treatment status and randomization strata for each sample dispensary.
- *pvalues_intermediate_outcomes*: List of p-values from the analysis of all intermediate outcomes, for the purpose of adjustment for multiple hypotheses testing.

Do files

The .do file titled *analysis* independently reproduces graphs and tables for primary and secondary outcomes, for both Indonesia and Tanzania. With the exception of this .do file, all other .do files are organized by country, as described below.

Indonesia

- *0_in_main*: Master .do file that sets the username and directories and runs .do files *1_in* through *multiple_hypotheses_testing_in*.
- *1_in*: Prepares a village-level dataset of average outcomes at baseline for primary and secondary outcomes, so that baseline village-level averages may be included in the final regression analyses as a robustness check. Also prepares measures of baseline health facility quality, to be used in the subgroup analysis.

- *2_in*: Cleans the data (including encoding missing data) needed to construct primary outcomes, constructs primary outcomes, and conducts the analysis, including robustness checks, subgroup analyses and detectable differences.
- *3_in*: Cleans the data needed to construct secondary outcomes, constructs secondary outcomes, and conducts the analysis, including robustness checks.
- *4_in*: For intermediate outcomes pertaining to household data. Cleans the data needed to construct intermediate outcomes, constructs intermediate outcomes, and conducts the analysis.
- *5_in*: For intermediate outcomes pertaining to facility data. Cleans the data needed to construct intermediate outcomes, constructs intermediate outcomes, and conducts the analysis.
- *6_in*: Cleans and analyses data on household awareness of potential activities to improve healthcare in their village.
- *multiple_hypotheses_testing_in*: Adjusts intermediate outcome p-values for multiple hypotheses testing.

Tanzania

- *0_tz_main*: Master .do file that sets the username and directories and runs .do files *1_tz* through *multiple_hypotheses_testing_tz*, as listed below.
- *1_tz*: Prepares a village-level dataset of average outcomes at baseline for primary and secondary outcomes, so that baseline village-level averages may be included in the final regression analyses as a robustness check. Also prepares measures of baseline health facility quality, to be used in the subgroup analysis.
- *2_tz*: Cleans the data (including encoding missing data) needed to construct primary outcomes, constructs primary outcomes, and conducts the analysis, including robustness checks, subgroup analyses and detectable differences.
- *3_tz*: Cleans the data needed to construct secondary outcomes, constructs secondary outcomes, and conducts the analysis, including robustness checks.
- *4_tz*: For intermediate outcomes pertaining to household data. Cleans the data needed to construct intermediate outcomes, constructs intermediate outcomes, and conducts the analysis.
- *5_tz* and *6_tz*: For intermediate outcomes pertaining to facility data. Cleans the data needed to construct intermediate outcomes, constructs intermediate outcomes, and conducts the analysis. *5_tz* analyzes facility-level intermediate outcomes, i.e. for a sample size of 153 facilities. *6_tz* analyzes village-level intermediate outcomes, for a sample size of 200 villages.
- *7_tz*: Cleans and analyses data on household awareness of potential activities to improve healthcare in their village.
- *multiple_hypotheses_testing_tz*: Adjusts intermediate outcome p-values for multiple hypotheses testing.

General instructions for replication

- Datasets and tables/graphs produced by the .do files are saved in the "Data" and "Analysis" folders, respectively.
- All .do files were created using Stata 15 and last modified in Stata 16.
- To recreate the tables and graphs for the main outcomes (i.e. primary and secondary), it is sufficient to independently run the .do file *analysis*. Prior to running this .do file, set your username and directory in the .do file.
- Running the master .do files, *0_main_in* and *0_main_tz*, will recreate all our analyses. This includes primary, secondary and intermediate outcomes, community awareness outcomes, subgroup analyses, robustness checks, detectable differences and adjustment of p-values for multiple hypotheses testing. Prior to running the master .do files, set your username and directory in the .do files.

Notes on data collection

- Pre-intervention data were collected between February and June 2015 in Indonesia, and March and July 2015 in Tanzania.
- Post-intervention data were collected between November 2017 and January 2018 in Indonesia, and May and August 2018 in Tanzania.
- In Indonesia, two sample villages were reassigned to different *puskesmas* between the pre- and post-intervention surveys. Pre-intervention, they fell under the jurisdiction of one *puskesmas*, but were later reassigned to a different *puskesmas*. In these cases, post-intervention data was collected from the *puskesmas* that was responsible for healthcare in the sample villages at the time of the intervention.
- In Tanzania, post-intervention data could not be collected from one dispensary because it had closed and did not offer any maternal or newborn health services over the duration of post-intervention data collection. Further, the post-intervention survey was only partially completed at one other dispensary.