Documentation file for Do-files and Data

"Leveraging text messages to improve community worker performance: Evidence from child protection in Sierra Leone"

Ryan Sheely[1]
Akshay Dixit[2]

Folder organization

- ➔ analysis
- ➔ data
- ➔ do files
- ➔ questionnaires

Note: The folder labelled "analysis" is empty and meant to store output produced by the .do files.

Questionnaires

The "questionnaires" folder contains four questionnaires in English, all in Excel format. These were used for the post-intervention (endline) data collection.

- endline-adult-final: Questionnaire administered to the adult respondents in the sample.
- endline-chief-final: Questionnaire administered to the sample village chiefs.
- endline-child-final: Questionnaire administered to the child respondents in the sample.
- endline-fp-final: Questionnaire administered to the Child Protection Focal Persons (CPFPs).

Data

Datasets listed below were deidentified. Personally identifying information, and combinations thereof, were removed to protect the identities of individuals.

- baseline_master: Pre-intervention (baseline) data from sample villages in comma delimited (.csv) as well as Stata dataset (.dta) formats. This includes data from the adult, chief, child and CPFP surveys.
- endline_master: Post-intervention (endline) data from sample villages in comma delimited (.csv) as well as Stata dataset (.dta) formats. This includes data from the adult, chief, child and CPFP surveys.
- pvalues: List of p-values from the analysis of all main and secondary outcomes, for the purpose of adjustment for multiple hypotheses testing.

[1] Mercy Corps, email: rsheely@mercycorps.org
[2] Harvard University, email: akshay_dixit@g.harvard.edu

Do files

- balance_check: Checks for balance between treatment and control groups on endline covariates, as well as on baseline outcomes.
- master: Master .do file that sets the username and directories and runs all .do files listed below.
- 0_organize: Separates the endline dataset by respondent; splitting CPFP survey data from adult, chief and child survey data.
- 1_baseline_social_capital: Prepares baseline outcome for heterogeneous treatment effects analysis by social capital
- 2_main_outcomes: Cleans the data (including encoding missing data) needed to construct the main outcomes, constructs the main outcomes, and conducts the analysis, including computing the ITT and TOT estimates.
- 3_cpfp_involvement: Cleans data and conducts the analysis of outcomes measuring the involvement of CPFPs in child protection activities.
- 4_other_involvement: Cleans data and conducts the analysis of outcomes measuring the involvement of various child protection actors in child protection activities.
- 5_child_protection: Cleans data and conducts the analysis of child protection outcomes.
- 6_multiple_hypotheses_testing: Adjusts p-values for multiple hypotheses testing.
- 7_cpfp_self_involvement: Cleans data and conducts the analysis of outcomes measuring the involvement of CPFPs in child protection activities, as reported by CPFPs themselves.
- 8_cpfp_other_involvement: Cleans data and conducts the analysis of outcomes measuring the involvement of various child protection actors in child protection activities, as reported by CPFPs.

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.
- Set your username and directory in the .do files prior to running them.
- Running the master .do file will recreate all our analyses, except for balance checks. This includes main outcomes, secondary outcomes, heterogeneous treatment effects analyses, analyses based on the CPFP data, and adjustment of p-values for multiple hypotheses testing.