

Documentation File for Data and Codes for the paper:

Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda

Martina Björkman Nyqvist, Andrea Guariso, Jakob Svensson, David Yanagizawa-Drott.

Introduction

Below we provide information on the files needed to reproduce the results from the paper *“Reducing Child Mortality in the Last Mile: Experimental Evidence on Community Health Promoters in Uganda”*. All variables in the datasets are named and labelled so to clarify their meaning.

Datasets

Overall, there are 13 datasets associated with this study. Of these, 11 datasets were constructed by the research team through original data collection. Other intermediate and final datasets are produced by running the code (see *Code* section). Below we provide a short description of the datasets.

1. *AEJ2018_HHmain.dta*: the dataset contains information at the household level. Each line refers to a different household. Answers were provided by the main respondent in the HH.
2. *AEJ2018_HHroster_u5*: the dataset contains the household roster for household members under 5 years of age. Each line refers to a different household member under 5 years of age. Answers were given by the main respondent in the HH.
3. *AEJ2018_HHroster_5to17*: the dataset contains the household roster for household members between 5 and 17 years of age. Each line refers to a different household member between 5 and 17 years of age. Answers were given by the main respondent in the HH.
4. *AEJ2018_HHroster_ab18*: the dataset contains the household roster for household members above 18 years of age. Each line refers to a different household member above 18 years of age. Answers were given by the main respondent in the HH.
5. *AEJ2018_anthro*: the dataset contains anthropometric measures of children under 5 years of age, as well as hemoglobin levels. Each line refers to a different child under 5 years of age. The measures were collected by specifically trained enumerators for each child under 5 living in the household.

6. *AEJ2018_child_mortality*: the dataset contains information on child mortality. The dataset includes both children that were alive at the time of the survey and children who died in the previous years. Each line refers to a different child (the variable “died” indicates children who died). Answers were given by the main respondent in the HH.
7. *AEJ2018_antenatal*: the dataset contains information on pregnant women in the household. Each line refers to a different pregnant woman. Answers were given by the pregnant women, whenever possible, or in alternative by the main respondent in the HH.
8. *AEJ2018_postnatal*: the dataset contains information on women in the household who delivered in the previous 12 months. Each line refers to a different woman who delivered in the previous 12 months. Answers were given by the women who delivered in the previous 12 months, whenever possible, or in alternative by the main respondent in the HH.
9. *AEJ2018_diarrhea*: the dataset contains a set of variables concerning diarrhea among children under 5 years of age living in the household. The questions were asked to the main respondent for each child under 5 living in the household. Each line refers to a different child under 5.
10. *AEJ2018_malaria*: the dataset contains a set of variables concerning malaria among children under 5 years of age living in the household. The questions were asked to the main respondent for each child under 5 living in the household. Each line refers to a different child under 5.
11. *AEJ2018_fortified*: the dataset contains a set of variables concerning fortified foods. The questions were asked to the main respondent for each child under 5 living in the household. Each line refers to a different child under 5.

There are two key identifiers that can be used to match observations across datasets:

- “hhid” is the household-specific unique identifier;
- “childID” is the identifier for children under 5 within each household

Two additional datasets have been generated using administrative records from the partner organizations that implemented the program.

12. *AEJ2018_sales_byCHP*: the dataset records the earnings of the Community Health Promoters (CHPs) in 2013. Each line refers to a different CHP.
13. *AEJ2018_CHPpurchases2013*: the dataset records the 2013 sales, divided by category of products. Each line is a different category of products sold by the CHPs.

Code

To simplify the replication, the codes have been divided according to the specific Table and Figure that they generate. The *Master* do file can be used to run all code at once, though each file can be run separately as well. In order to generate some Tables, a preliminary do file needs to be run to set up the dataset. The auxiliary do file is automatically launched by the corresponding do file, though it can also be run separately.

Stata Version

The data and code are compatible with Stata versions 14 and up.