README and Guidance for Reproducibility Package

November 2024

Effects of a community-driven water, sanitation, and hygiene intervention on diarrhea, child growth, and local institutions: a cluster-randomized controlled trial in rural Democratic Republic of Congo

by John P Quattrochi, Kevin Croke, Caleb Dohou, Luca Stanus Ghib, Yannick Lokaya, Aidan Coville, Eric Mvukiyehe (November 2024)

Overview

This reproducibility package includes all code and data necessary to reproduce the tables and figures presented in the paper. Minor differences in table formatting may appear due to manual adjustments. Within the code folder, you'll find Stata do files, programs, and ado files that handle the data preparation, analysis, and export of tables and figures. To replicate the analysis, simply update the working directory path in the master Stata do file (`Project_MasterDofile.do`) to point to the project root (the location of the "Reproducibility package" folder on your machine) and run the script. This script will automatically install load required packages/programs and execute the analysis. Output tables and figures are saved directly into their designated sub-folders, and two compiled `.docx` files in the root folder gathers all tables and figures in one document for easy reference. There are 10 tables. The replicator should expect the code to run for about 22 seconds.

Data Availability and Provenance Statements

The data for this project are confidential, but may be obtained with Data Use Agreements with the Development Impact Evaluation Department of the World Bank. Researchers interested in access to the data may contact Aidan Coville at acoville@worldbank.org. It can take a few months to negotiate data use agreements and gain access to the data. The authors will assist with any reasonable replication attempts for two years following publication.

Statement about Rights

- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

- I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permission are documented in the LICENSE.txt file.

License for Data

See LICENSE.txt for details.

Summary of Availability

- Data will be available in the World Bank Microdata Library in the future. In the meantime, data used in this paper will be available to individuals and organizations upon request.

Folder Structure

This reproducibility package is organized into the following folders:

- +--Reproducibility_Package
- +------Manuscript
- +------Health_Paper
- +------------Dofile
- +------------------Master_Data
- +------------------Household
- +------------------Village_leadership
- +------------------Anthropometric
- +------------------Water_point_mapping
- +------------------Household_water_testing
- +------------Data
- +------------------Master_Data
- +------------------Raw
- +------------------Intermediate
- +------------------Final
- +------------Documentation
- +------------------02_igrowup
- +-------------------------99_igrowup_ados
- +-------------------------anthro_z_scores
- +-------------------------macros_reftables
- +------------Analysis
- +------------------Tables

Details on each Data Source

All data provided here have been commissioned by our team and collected as part of the research we have conducted. Only two datasets are a combination of secondary and primary data collected by health zone and transmitted to us for randomization and sampling purposes. No personal identifying information is provided.

Data.Name	Data.Files	Location	Provided
Master_Data folder			
Intermediary list of randomized villages	rando_fulldata.dta	Health_Paper/Data/Master_Data/	TRUE
Final list of randomized villages (obtained after randomly dropping a certain number of villages from the intermediary list of randomized villages following a specific rule)	village_list_health_paper.dta	Health_Paper/Data/Master_Data/	TRUE
Raw data folder			
Raw data: Water quality test of water stored in household for drinking purposes (collected between Nov 2022- Feb 2023)	HH_Water_Test_Endline_raw_health_paper.dta	Health_Paper/Data/Raw/	TRUE
Raw data: Water point mapping data and water quality data at source (collected between Nov 2022- Feb 2023)	WP_Mapping_Endline_raw_health_paper.dta	Health_Paper/Data/Raw/	TRUE
Raw data: Anthropometrics data “height, weight, age and sex” for children under 5 in sampled	Anthro_Endline_raw_health_paper.dta	Health_Paper/Data/Raw/	TRUE

household (collected between Nov 2022- Feb 2023)			
Raw data: Village leadership data including information on WASH organization and committee activities (collected between Nov 2022- Feb 2023)	Leaders_Endline_raw_health_paper.dta	Health_Paper/Data/Raw/	TRUE
Raw data: Household data on water use, child health, knowledge and self-reported behavior, satisfaction with committee activities and WASH matters (collected between Nov 2022- Feb 2023)	Household_Endline_raw_health_paper.dta	Health_Paper/Data/Raw/	TRUE
Intermediate data folder			
Clean data: Water quality test of water stored in household for drinking purposes	HH_Water_Test_Endline_clean_health_paper.dta	Health_Paper/Data/Intermediate/	TRUE
Clean data: Water point mapping data and water quality data at source	WP_Mapping_Endline_clean_health_paper.dta	Health_Paper/Data/Intermediate/	TRUE
Clean data: Anthropometrics data "height,	Anthro_Endline_clean_health_paper.dta	Health_Paper/Data/Intermedi	TRUE

weight, age and sex" for children under 5 in sampled household		ate/	
Clean data: Village leadership data including information on WASH organization and committee activities	Leaders_Endline_clean_health_paper.dta	Health_Paper/Data/Intermediate/	TRUE
Clean data: Household data on water use, child health, knowledge and self-reported behavior, satisfaction with committee activities and WASH matters	Household_Endline_clean_health_paper.dta	Health_Paper/Data/Intermediate/	TRUE
Final data folder			
Analysis data: Water quality test of water stored in household for drinking purposes	HH_Water_Test_Endline_constructed_health_paper.dta	Health_Paper/Data/Final/	TRUE
Analysis data: Water point mapping data and water quality data at source	WP_Mapping_Endline_constructed_health_paper.dta	Health_Paper/Data/Final/	TRUE
Analysis data: Anthropometrics data including z-scores for children under 5	Anthro_z_st_constructed_health_paper.dta	Health_Paper/Data/Final/	TRUE
Analysis data: Village leadership	Leaders_Endline_constructed_health_paper.dta	Health_Paper/Data/	TRUE

data including WASH institutions and committee activities (collected between Nov 2022- Feb 2023)		Final/	
Analysis data: Household data on water use, child health, knowledge and self-reported behavior, satisfaction with committee activities and WASH matters (collected between Nov 2022- Feb 2023)	Household_Endline_constructed_health_paper.dta	Health_Paper/Data/Final/	TRUE
Analysis data: Household child level data including child health information	Chlid_roster_Endline_health_paper.dta	Health_Paper/Data/Final/	TRUE
Analysis data: Household level structured observations on knowledge and behavior	Structured_obs_Endline_hh_level_health_paper.dta	Health_Paper/Data/Final/	TRUE

The Stata data files (.dta) used to support the findings of this study were collected by the authors, and will be available in the World Bank Microdata Library.

Code for data cleaning, construction, and analysis is provided as part of the replication package. It is available in "Health_Paper/Dofiles" folder for review.

Computational requirements

Software Requirements

The statistical software Stata was essentially used for data processing and analysis. For smooth execution in Stata, the data processing and analysis required the installation of packages, ado files and a user-written program. The list of packages, ado files and a user-written program is also provided below.

- Stata (code was last run with version 16.0)
  - `estout` (as of 2023-02-12)
  - `ietoolkit` (as of 2024-02-12)
  - `winsor2` (as of 2020-11-25)
  - `WMEANEFFECTS` (user-written program)
  - `igrowup_standard` (ado as of 2007-23-05)

Controlled Randomness

- No Pseudo random generator is used in the analysis described here.

Memory, Runtime, Storage Requirements

Summary

Approximate time needed to reproduce the analyses on a standard (2024) desktop machine: 22 seconds

Approximate storage space needed: 302 MB

Details

The code was last run on a **MacBook Pro, Apple M1 Pro, macOS Sequoia Version 15.1, 16 GB of Memory with 276.6GB of free space.** Computation took **30 seconds.**

Description of programs/code

- The code right in "Health_Paper" and named "`Project_MasterDofile.do`" run all code and program as part of this analysis and produces all outputs.
- "Health_Paper/Dofiles/Master_Data" folder should include:
  - a code named "`village_list_health_paper.do`" that produces an intermediary and final village list. The final list is later merged with all survey datasets. Intermediate and final village lists are used to produce "table_s1_hp.xlsx" using .do file "table_s1_health_paper.do" from folder "Health_Paper/Dofiles/Tables".

  - Inside folder "Programs" a program named "`index_construction_program.do`" that is used to generate key indexes such as the WASH institutions index and the WASH governance perception index, used as part of this research. Indices are calculated by re-scaling each variable in each index so that higher values imply better outcomes, then standardizing relative to the control group, following Kling, Liebman, and Katz (2007)

- "Health_Paper/Documentation/02_igrowup/99_igrowup_ados" folder includes an ado file named "`igrowup_standard.ado`" that is used to compute z-scores within the data construction anthropometrics code. The macro calculates z-scores for the eight anthropometric indicators including weight-for-age, length/height-for-age, weight-for-length/height among others, and based on the WHO Child Growth Standards[1]. In this macro, all available (non-missing and non-flagged) z-score values are used for each indicator-specific prevalence estimation (standard analysis). Further information could be found in folder "Health_Paper/Documentation"

- "Health_Paper/Dofiles/Household" folder includes:

  - A deidentification code that also extracts only the necessary household data for the paper. The code is named "`01_hh_deid_health_paper.do`". This code is referenced in the master do-file but is set as a comment for it not to run since the input data here includes personal identifiable data. The do-file outputs the deidentified raw data used for all data processing and analysis.

  - A cleaning code named "`02_hh_cleaning_health_paper.do`" that performs cleaning and outputs a clean version of the household data.

  - A code named "`03_hh_construct_health_paper.do`" that uses the clean household data as input, performs variable construction and outputs the final analysis data.

- "Health_Paper/Dofiles/Village_leadership" folder includes:

  - A deidentification code that also extracts only the necessary variables from Village leadership data for the analysis. The code is named "`01_vl_deid_health_paper.do`". This code is referenced in the master do-file but is set as a comment for it not to run since the input data here includes personal identifiable data. The do-file outputs the deidentified raw data used for all data processing and analysis.

  - A cleaning code named "`02_vl_cleaning_health_paper.do`" that performs data cleaning and outputs a clean version of the Village leadership data.

---

[1] Detailed citations can be found at the end of this document in the section Reference.

  - A code named "03_vl_construct_health_paper.do" that uses the clean village leadership data as input, performs variable construction and outputs the final analysis data.

- "Health_Paper/Dofiles/Anthropometric" folder includes:
  - A deidentification code that also extracts only the necessary variables from Anthropometrics data for the analysis. The code is named "01_anthro_deid_health_paper.do". This code is referenced in the master do-file but is set as a comment for it not to run since the input data here includes personal identifiable data. The do-file outputs the deidentified raw data used for all data processing and analysis.
  - A cleaning code named "02_anthro_cleaning_health_paper.do" that performs cleaning and outputs a clean version of the Anthropometrics data.
  - A code named "03_anthro_construct_health_paper.do" that uses the clean anthropometrics data as input, performs variables construction and outputs the final analysis data including variables such as z-scores, among others.

- "Health_Paper/Dofiles/Water_point_mapping" folder includes:
  - A deidentification code that also extracts only the necessary variables from Water point mapping data for the analysis. The code is named "01_wp_deid_health_paper.do". This code is referenced in the master do-file but is set as a comment for it not to run since the input data here includes personal identifiable data. The do-file outputs the deidentified raw data used for all data processing and analysis.
  - A cleaning code named "02_wp_cleaning_health_paper.do" that performs cleaning and outputs a clean version of the Water point mapping data.
  - A code named "03_wp_construct_health_paper.do" that uses the clean water point mapping data as input, performs variable construction and outputs the final analysis data.

- "Health_Paper/Dofiles/Household_water_testing" folder includes:
  - A deidentification code that also extracts only the necessary variables from household water testing data for the analysis. The code is named "01_hhw_deid_health_paper.do". This code is referenced in the master do-file but is set as a comment for it not to run since the input data here includes personal identifiable data. The do-file outputs the deidentified raw data used for all data processing and analysis.

- A cleaning code named "02_hhw_cleaning_health_paper.do" that performs cleaning and outputs a clean version of the Household water testing data.
- A code named "03_hhw_construct_health_paper.do" that uses the clean household water testing data as input, performs variable construction and outputs the final analysis data.

- "Health_Paper/Dofiles/Tables" folder includes:
  - The code named "table_2_health_paper.do" takes the household analysis data as input and produces "Table 2. Household and respondent characteristics by intervention group, at 3.5-year follow-up" that could be found in "Health_Paper/Analysis/table_2_hp.xslx".
  - The code named "table_3_health_paper.do" takes the household, anthropometrics z-score, and village leadership analysis data as input and produces "Table 3. Intervention effects on primary outcomes: diarrhea, length-for-age, and WASH institutions", that could be found in "Health_Paper/Analysis/table_3_hp.xslx".
  - The code named "table_4_health_paper.do" takes the household data, household structured observations data, water point mapping data, household water testing data, anthropometrics z-scores data, household children level data as input and produces "Table 4. Intervention effects on secondary outcomes related to water, sanitation, and hygiene", that could be found in "Health_Paper/Analysis/table_4_hp.xslx".
  - The code named "table_s1_health_paper.do" takes the intermediate randomization data and outputs "Table S1. Randomization strata", that could be found in "Health_Paper/Analysis/table_s1_hp.xslx".
  - The code named "table_s2_health_paper.do" takes the village leadership data as input and produces "Table S2. Intervention effects on WASH institutions index and index sub-components", that could be found in "Health_Paper/Analysis/table_s2_hp.xslx".
  - The code named "table_s3_health_paper.do" takes the household data, household structured observations data, water point mapping data, household water testing data, anthropometrics z-scores data, household children level data as input and produces "Table S3. Intervention effects on all secondary outcomes, including index sub-components", that could be found in "Health_Paper/Analysis/table_s3_hp.xslx".
  - The code named "table_s4_health_paper.do" takes the anthropometrics z-scores data, household children level data, and village leadership data as input and produces "Table S4. Intervention effects on all primary outcomes, by

province (pre-specified)", that could be found in "Health_Paper/Analysis/table_s4_hp.xslx".

- The code named "`table_s5_health_paper.do`" takes the anthropometrics z-scores data, household children level data, and village leadership data as input and produces "Table S5. Intervention effects on all primary outcomes, province-by-intervention interaction models", that could be found in "Health_Paper/Analysis/table_s5_hp.xslx".
- The code named "`table_s6_health_paper.do`" takes the anthropometrics z-scores data, household children level data as input and produces "Table S6. Intervention effects on diarrhea and length-for-age z score, separately by sex (pre-specified)", that could be found in "Health_Paper/Analysis/table_s6_hp.xslx".
- The code named "`table_s7_health_paper.do`" takes the anthropometrics z-scores data, household children level data, and village leadership data as input and produces "Table S7. Intervention effects on diarrhea and length-for-age z score, sex-by-intervention interaction models", that could be found in "Health_Paper/Analysis/table_s7_hp.xslx".

License for Code

Instructions to Replicators

- Download the folder "Reproducibility package" to your preferred location
- Obtain the data once it's published in the World Bank Microdata Library or reach out to the project data contact for earlier access.
- Within the "reproducibility package" folder, you'll find a folder "Health_Paper" that includes all the necessary resources.
- Open the code named "`Project_MasterDofile.do`" located right in the "Health_Paper" folder
- Go to the subsection "**** Folder paths necessary for running this reproducibility package" and adjust the relevant default path under that section: Add the path of the downloaded folder to the folder "Health_Paper".
- Make sure you are connected to the internet to allow for installation of all packages.

- Once the previous steps successfully applied, you can run the "Project_MasterDofile.do" that will produce all relevant outputs including the tables used as part of the paper.

List of tables and programs

The provided code reproduces:

- All numbers provided in text in the paper
- All tables and figures in the paper
- Selected tables and figures in the paper, as explained and justified below.

Exhibit name	Output filename	Script	Note
Table 1. The Healthy Village and Schools Program's Nine Steps			Produced manually and include the steps to complete the intervention provided as part of the impact evaluation.
Table 2. Household and respondent characteristics by intervention group, at 3.5-year follow-up	table_2_hp.xlsx	table_2_health_paper .do	Found in Health_Paper/Analysis/ Tables
Table 3. Intervention effects on primary outcomes: diarrhea, length-for-age, and WASH institutions	table_3_hp.xlsx	table_3_health_paper .do	Found in Health_Paper/Analysis/ Tables
Table 4. Intervention effects on secondary outcomes related to water, sanitation, and hygiene	table_4_hp.xlsx	table_4_health_paper .do	Health_Paper/Analysis/ Tables
Table S1. Randomization strata	table_s1_hp.xlsx	table_s1_health_paper .do	Health_Paper/Analysis/ Tables
Table S2. Intervention effects on WASH institutions index and index sub-components	table_s2_hp.xlsx	table_s2_health_paper .do	Health_Paper/Analysis/ Tables
Table S3. Intervention	table_s3_hp.xlsx	table_s3_health_paper	Health_Paper/Analysis/

Exhibit name	Output filename	Script	Note
effects on all secondary outcomes, including index sub-components		.do	Tables
Table S4. Intervention effects on all primary outcomes, separately by province (pre-specified)	table_s4_hp.xlsx	table_s4_health_paper .do	Health_Paper/Analysis/ Tables
Table S5. Intervention effects on all primary outcomes, province-by-intervention interaction models	table_s5_hp.csv	table_s5_health_paper .do	Health_Paper/Analysis/ Tables
Table S6. Intervention effects on diarrhea and length-for-age z score, separately by sex (pre-specified)	table_s6_hp.xlsx	table_s6_health_paper .do	Health_Paper/Analysis/ Tables
Table S7. Intervention effects on diarrhea and length-for-age z score, sex-by-intervention interaction models	table_s7_hp.csv	table_s7_health_paper .do	Health_Paper/Analysis/ Tables
Table S8. Variable definitions			Produced manually. It provides details on variables used as part of the analysis and their definition
S9. CONSORT 2010 checklist of information to include when reporting a randomised trial			Produced manually. It includes checklist when reporting a randomized trial

References

Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. Econometrica, 75(1), 83-119.

WHO Multicentre Growth Reference Study Group (2006). WHO Child Growth Standards: Length/height-for-age, weight-for-age, weight-for-length, weight-for-height and body mass index-for-age: Methods and development. Geneva: World Health Organization; pp 312.

(available on the web site: http://www.who.int/childgrowth/publications/en/)

WHO Multicentre Growth Reference Study Group (2007). WHO Child Growth Standards: Head circumference-for-age, arm circumference-for-age, triceps skinfold-for-age and subscapular skinfold-for-age: Methods and development. Geneva: World Health Organization

---

Acknowledgements

Some content on this page was copied from Hindawi. Other content was adapted from Fort (2016), Supplementary data, with the author's permission.