Plan Overview

A Data Management Plan created using DMP Tool

DMP ID: https://doi.org/10.48321/D1C36P

Title: See it for Yourself: E. Coli Test Kits and Ceramic Water Filter Compliance in the Kibera Informal Settlement

Creator: Hamid Oskorouchi - **ORCID:** 0000-0003-1241-9645

Affiliation: University of Hohenheim (uni-hohenheim.de)

Principal Investigator: Marta Parigi, Hamid Reza Oskorouchi, Alfonso Sousa-Poza

Data Manager: Marta Parigi, Hamid Reza Oskorouchi

Funder: German Federal Ministry For Education And Research (bmbf)

Grant: 01DG21024

Template: Digital Curation Centre

Project abstract:

Bacterial contamination of drinking water is a major cause of death, especially among poorer children. In settings where water infrastructure is inadequate or even non-existent, policy makers have often relied on the provision of point-of-use low-tech water purification methods. Despite the extant literature shows the latter being extremely cost-effective in reducing waterborne diseases and even mortality, compliance with these technologies remains low. In this study, by means of a clustered randomized controlled trial, we will test if the provision of easy-to-use and -to-interpret drinking water test kits for coliforms can increase usage of ceramic water filters (CWF) distributed freely to 1,008 households living in the informal settlement of Kibera, Nairobi. For this analysis one of our primary outcomes consists into an objective measure of the amount of filtered water collected by means of a domestic water meter installed inside each CWF. We will also estimate the causal effect of our intervention on a series of secondary health and non-health outcomes (including child diarrhea and wasting). Finally, we will investigate if the provision of tests kits

produces heterogenous effects depending on the outcome of the test (positive vs. negative) and the type of water tested (raw vs. filtered).

Start date: 05-31-2021

End date: 06-01-2024

Last modified: 07-08-2024

Copyright information:

See it for Yourself: E. Coli Test Kits and Ceramic Water Filter Compliance in the Kibera Informal Settlement

Data Collection

What data will you collect or create?

With this project, we plan the collection of a three-wave household-level longitudinal dataset (listing: N~1.800, baseline: N~1,000 and endline: N~1,000). The data collection will be performed using digital devices (tablets) and the SurveyCTO platform and servers.

The sample includes households residing in the Kibera informal settlement, Nairobi, Kenya, for which a respondent at least 18 years old is willing to participate in the survey after having given written consent. Collection of child anthropometrics will require additional written consent from at least one of the child's parents(or legal guardians).

The data will include information about household members' demographics, education, health, income, wealth, subjective wellbeing, child anthropometrics (weight, height, and MUAC), and a series of confidential identifying information (name and surname of the household members, respondent's primary and secondary telephone numbers, and the dwelling's GPS coordinates) needed to retrack the households in subsequent waves of the longitudinal data collection.

How will the data be collected or created?

The survey data will be collected by means of on-site household visits by local enumerators.

Two main files will be created:

·   Listing data (about 1,800 households), a dataset to be used primarily for ex-ante power analyses

·   Baseline and Endline data, a longitudinal dataset nested within the listing data that will include around 1,000 specially selected listing data observations.

From the dataset mentioned above, the authors will create other files containing datasets reshaped at different aggregation levels (e.g., mothers, children, etc...). Variables for which value have been imputed will be clearly labeled, and the precise imputation procedure will be detailed in the data codebook.

Documentation and Metadata

What documentation and metadata will accompany the data?

The data will be accompanied by the original SurveyCTO questionnaire (in its totality or only in part) and a codebook additionally describing the authors' manipulations after data collection (e.g., index construction, variable transformations). The data will be shared in Stata format (dta) and on request in formats that can be read using open-source software (e.g., Rda, or csv).

Ethics and Legal Compliance

How will you manage any ethical issues?

This study has gained consent from the survey participants to preserve and share anonymized data for research purposes only, including future unplanned projects. After each data collection activity, the data will be permanently deleted from the SurveyCTO server, and saved on the secured hard-disks of the PIs' business laptops (if pseudoanonymized) or on offline password protected media (if containing confidential identifying information such as names and surnames, telephone numbers, or GPS coordinates).

Ethical approval for this study was obtained from both the Ethics Committee of the University of Hohenheim, Stuttgart, Germany (PI's home institution) (3 December 2021 and 24 February 2022) and the Maseno University Ethics Committee (28 February 2022). All authors comply strongly with the rules of good scientific practice as defined by the German Research Foundation (DFG) in "Safeguarding Good Scientific Practice."

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

The data collected in this project are the property of the University of Hohenheim and can be exclusively used by the project PIs (together with external coauthors). Anonymized versions of the data can be shared only for purposes of peer review or replication in approved research intended for publication. [jca1] Under no circumstances can nonanonymized data be shared with any third party.

Storage and Backup

How will the data be stored and backed up during the research?

Confidential data:

· Deletion from the SurveyCTO server within a month of the end of each data collection activity.

· Back-up on offline password protected media.

Pseudoanonymized data:

· Created by the authors by deleting any identifying information (including names and surnames, telephone numbers, and GPS coordinates).

· Back-up on PIs secured business laptops.

All data, questionnaires and do-files will be backed-up on password protected CDs (at least two copies) after each data collection activity. The CDs will be stored in a locked drawer in the PIs offices at the University of Hohenheim, Stuttgart, Germany.

· Person responsible for back-up of confidential data: H.R. Oskorouchi

· Person responsible for pseudoanonymization and anonymization procedures: H. R. Oskorouchi

· Persons responsible for back-up of pseudoanonymized data: all authors.

How will you manage access and security?

The risk to data security is limited by the following practices: (i) all confidential information will be stored offline in password protected media, and ii) the authors will not publicly share the complete dataset. (iii) Although all study authors will have easy access to the pseudoanonymized version of the data, only the corresponding author will handle the confidential data. (iv) During the programming of the SurveyCTO questionnaire, the programmer of the Nairobi-based data collection firm handling the field work will have access to the household GPS coordinates but to NO OTHER identifying information except for randomly assigned household or building identifiers. (v) Coauthors external to this this project will have access to pseudoanonymized versions of the data upon signing a binding nondisclosure statement, and (vi) only the project PIs have the necessary permissions to download the data from the secure SurveyCTO servers.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Which data are of long-term value and should be retained, shared, and/or preserved[jca1] ?

· Data to be used for future exploratory analyses and/or descriptive studies.

· Data preserved under the modalities outlined above.

No data need to be destroyed in the future.

What is the long-term preservation plan for the dataset?

Other than possible publication of the anonymized data for replication purposes, no plan or funding currently exists to archive the full dataset into a repository.

Data Sharing

How will you share the data?

Data for replication purposes will be publicly shared on the webpage of the journal in which the project's articles will be published. Alternatively, if the journal offers no medium for direct sharing, the dataset will be published on Github.

Are any restrictions on data sharing required?

Because Kibera is a relatively small subcounty of Nairobi, access to the complete household roster would pose confidentiality risks, so the complete dataset cannot be shared for fear of household identification.

Responsibilities and Resources

Who will be responsible for data management?

All PIs will be responsible for implementing the data management plan (DMP) and informing prospective third-party users (e.g., reviewers and external coauthors) of its constraints.

What resources will you require to deliver your plan?

DMP implementation requires no special resources other than those offered by the PIs' home institution to all its scientific employees.