Tanzania - Water Supply & Expansion

Report generated on: March 8, 2018

Visit our data catalog at: https://data.mcc.gov/evaluations/index.php

Overview

Identification

COUNTRY
Tanzania

EVALUATION TITLE
Water Supply & Expansion

EVALUATION TYPE
Independent Impact Evaluation

ID NUMBER
DDI-MCC-TZA-WASH-IE-ML-SI-2017-v01.0

Version

VERSION DESCRIPTION
- v01: Edited, anonymous dataset for public distribution.

Overview

ABSTRACT
Social Impact (SI) has been contracted by MCC to carry out an impact evaluation (IE) of the Tanzania Water Sector Project. This IE examines the effect of the WSP interventions on outcomes such as water consumption, time spent collecting water, diarrheal illness, and investments in human and physical capital. Given that the impacts of the upgraded water infrastructure are expected to be diffuse across each city, identifying a counterfactual through experimental methods is not feasible. SI has designed a rigorous quasi-experimental impact evaluation, using generalized propensity score matching (GPSM). This design utilizes a continuous treatment approach. The evaluation is designed to be able to quantify the impacts of incremental changes in water supply.

SI supplements the main design with qualitative, geospatial, and administrative data, as well as additional statistical analyses using Instrumental Variable (IV) and Ordinary Least Squares (OLS) regression, to triangulate findings from the main impact analyses.

EVALUATION METHODOLOGY
Continous Treatment

UNITS OF ANALYSIS
Main analysis: households and individuals. (Additional analyses with water quality data are analyzed at the cluster, or enumeration area, level as well as at the water treatment plant).

KIND OF DATA
Sample survey data [ssd]

TOPICS

Topic	Vocabulary	URI
Water, Sanitation, and Hygiene	MCC Sector	

KEYWORDS
Water, Urban, Tanzania, Impact Evaluation, Generarlized Propensity Score Matching, Health, Productivity, Poverty, Infrastructure

Coverage

GEOGRAPHIC COVERAGE
Urban municipalities of the cities of Dar es Salaam (three municipalities including Ilala, Kinondoni, and Temeke) and Morogoro (Morogoro Urban municipality)

UNIVERSE
Midline data collection included a process study, a qualitative cognitive interviewing activity, a household survey, water quality testing, and direct measurement of water availability using pressure logging devices, as well as secondary data collection from utilities and other agencies in Tanzania.

The household surveys were administered to one respondent per household, with additional questions asked of each current household member (usual residents). The water quality tests were administered to up to two sources per cluster (either household tap, or other shared source in the cluster) as well as up to one source per cluster for point of consumption tests. Water quality tests were conducted on raw and treated water samples from water treatment plants (at the plant intakes and outlets, respectively). Data from pressure logging devices were collected from a sample of household respondents.

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
Social Impact, Inc.	

FUNDING

Name	Abbreviation	Role
Millennium Challenge Corporation	MCC	

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
Millennium Challenge Corporation	MCC		Review of Metadata
Social Impact	SI		Independent Evaluator (Produce Metadata)

DATE OF METADATA PRODUCTION
2017-01-20

DDI DOCUMENT VERSION
Version 1.0 (Original 2017-01-20): Original version of metadata for midline data collection activities

DDI DOCUMENT ID
DDI-MCC-TZA-WASH-IE-ML-SI-2017-v01.0

MCC Compact and Program

COMPACT OR THRESHOLD
Tanzania Compact

PROGRAM
In 2008, the United States Millennium Challenge Corporation (MCC) and the Government of Tanzania (GoT) entered into a Millennium Challenge Compact to facilitate poverty reduction through economic growth. This $698 million Compact included a $64.2 million Water Sector Project (WSP) implemented by the Millennium Challenge Account-Tanzania (MCA-T). The WSP aimed to improve the supply and quality of water in the two urban centers of Morogoro and Dar es Salaam through investments in water infrastructure. Dar es Salaam is the largest city and economic center of Tanzania, with a population of about 4.36 million. Morogoro is a region about 190 kilometers west of Dar es Salaam whose urban municipality has a population of about 316 thousand. Both cities' populations are growing rapidly. In the decade prior to the 2012 census, Dar es Salaam grew by an estimated 5.6 percent per year, while Morogoro grew by 3.3 percent per year. The Morogoro Water Supply Activity included upgrades and rehabilitation to two water treatment plants (WTP), Mafiga and Mambogo, increasing

water supply in the distribution network from 23 to 33 million liters per day (MLD), while also improving the quality of water produced. In Dar es Salaam, the Lower Ruvu Plant Expansion Activity aimed to expand the capacity of the Lower Ruvu WTP from 180 to 270 MLD. The interventions in Morogoro were put into operation in mid-2014. Although the interventions in Dar es Salaam were completed in 2015, they did not take effect until a new transmission main funded by the GoT was completed in March 2016.

MCC SECTOR

Water, Sanitation and Hygiene (WASH)

PROGRAM LOGIC

Overall, the MCC Tanzania Water Sector Project aims to increase investment in human and physical capital and to reduce the prevalence of water-related disease, in order to reduce poverty through economic growth. Specifically, the improvements to water infrastructure in the two cities are expected to result an increase in availability of water for households across the utilities' coverage areas; as increased volume of water flows through piped water system, water access (and potentially consumption) should increase. As a result of increased water access, the amount of time spent fetching water decreases, and household members can allocate this time to more productive activities, including income-generating activities (or, in the case of children, increased attendance at school). The interventions are also expected to have an effect on the quality of water accessible through the public distribution network, either directly through better-quality piped water, or indirectly through a substitution of piped water in place of water obtained from poorer quality sources. Increased access to clean, piped water, would be expected to result in lower prevalence of water-related diseases among adults and children. In turn, reduced illness, and reduced time spent caring for sick children, would permit an increase in the time allocated by adults to productive activities. In summary, the project logic posits that improvements in the quantity and reliability of the water supply will reduce the real and opportunity costs resulting from scarce or low-quality water, thus allowing households to allocate more time and resources to income-generating activities, promoting poverty reduction goals.

PROGRAM PARTICIPANTS

Program participants include groups targeted by the MCC Tanzania Water Sector Project, and included the entire populations of urban Dar es Salaam and Morogoro. Beneficiaries are considered to be all households (in Dar es Salaam and Morogoro) with connections to the improved water supply systems who experience an increase in incomes due to the project, especially those with connections to the public water distribution network. The parastatal water utilities in each city, including the Dar es Salaam Water and Sewerage Authority (DAWASA) and the Morogoro Urban Water and Sewerage Authority (MORUWASA), were the implementing partners to MCA-T, and manage the infrastructure after completion of the interventions. Surveys conducted as part of the impact evaluation therefore sampled from the full populations of each city (described in more detail in 'Study Population').

Sampling

Study Population

Midline data collection included a process study, a qualitative cognitive interviewing activity, a household survey, water quality testing, and direct measurement of water availability using pressure logging devices, as well as secondary data collection from utilities and other agencies in Tanzania. The household surveys were administered to one respondent per household, with additional questions asked of each current household member (usual residents). The water quality tests were administered to up to two sources per cluster (either household tap, or other shared source in the cluster) as well as up to one source per cluster for point of consumption tests. Water quality tests were conducted on raw and treated water samples from water treatment plants (at the plant intakes and outlets, respectively). Data from pressure logging devices were collected from a sample of household respondents.

Sampling Procedure

The household sampling procedures were completed in 2013 for the baseline phase using a two-stage stratified cluster sampling design. The baseline data collection firm was EDI.

Baseline Sampling Procedure: Households were sampled from both cities using a two-stage cluster sampling methodology, with stratification in Dar es Salaam by the current water supply to an area. Clusters were defined as census enumeration areas (EAs). The sample frame for clusters was an inventory of enumeration areas used for the 2012 census in Tanzania, obtained from the Tanzania National Bureau of Statistics (NBS). The required sample size was 5008 households (8 households from 626 clusters), split between the two cities evenly. In Morogoro, 313 clusters were randomly sampled from the master inventory. In Dar es Salaam, with the availability of water supply by ward, clusters were taken by stratified random sample, out of 5 strata corresponding to different levels of current water access through the public distribution network. Selection of clusters was done using a random number generator in Stata 12 software. After selecting 313 clusters in each city, maps were obtained from the NBS in order to guide EDI's listing teams. For each of the selected clusters, EDI's listing teams worked with local community representatives to generate a complete sample frame of households. From each cluster's household list, 8 households were randomly selected for the household survey from a unique random number table for each cluster; additional households from the list could be accessed in order to replace households as needed due to non-response.After the households were interviewed, a sub-set of eligible households - up to two per cluster - was selected for water quality testing. Following the household survey, the full household sample was included in three rounds of a follow-up survey administered by phone, by the EDI team.

Midline:The data collection firm at midline was CSR Group Africa. A tracking exercise was therefore implemented prior to the household survey, to locate households that had participated in the baseline survey. Where households from the baseline round could not be successfully located, had moved out of their original mtaa, or declined to participate in the midline survey, households were replaced in order to achieve the same sample size as in baseline. The replacement protocol prioritized households living in the same dwelling as any household who had moved, with the next most preferable option being a neighboring household (a contiguous neighbor to the original dwelling as much as possible). Neighboring households would be the first option for replacing households that had declined to participate. In total, 45 percent of households (1,122) in Morogoro had to be replaced along with 62 percent of households (1,553) in Dar es Salaam. An additional 14 households were added to the sample in total - these were households from surveyed clusters with piped connections that agreed to pressure logger installations; extra surveys were done with these households to obtain survey data corresponding to the pressure logging data. A total of 5,022 unique households were interviewed at midline.

For the cognitive interviewing activity, the sampling involved two stages for this activity: stratified random sampling of mtaa areas using baseline data, and purposive sampling of households based on recommendations of mtaa leaders. The mtaa areas selected are listed in Table A3. Areas were randomly selected from two strata in each city. The strata included piped and non-piped, defined by whether any household in that mtaa reported having a piped connection at baseline. Given the focus of the impact evaluation on improvements to the piped distribution network, greater emphasis was placed on piped households. Therefore, nine mitaa in each city were selected from the piped stratum, and the remaining three from the unpiped stratum..

Deviations from Sample Design

The household survey was envisioned as a panel of households, consistent across survey waves (e.g., baseline, midline, endline). Given the delays in project implementation, a two-year time period had elapsed between baseline and midline. Combined with the transience of many households in these dynamic urban settings, this led to a high degree of attrition

among the baseline household sample. Sampling procedures dictated that the top preference for a replacement was the new household living in the original dwelling, and where that was not feasible a contiguous neighbor was the next preferable option. This strategy was meant to include households that had comparable demographics and levels of access to water as those they replaced. It appears likely that the success rate of recruiting the new households in the original dwelling was low, thus leading to a group of replacement households that are less transient than those they replaced. Unfortunately, while records of the proportion of same-dwelling versus contiguous-neighbor replacements would allow us to answer this question definitively, such records were not kept during midline data collection.

Response Rate

Response rates for panel household survey in Dar es Salaam were 38%, and above 55% in Morogoro. Due to the high levels of attrition, a tracking exercise and replacement protocol were put in place to achieve the sample size.

Weighting

The same sampling weights as baseline are applied to the midline dataset. Therefore, unless otherwise specified, the midline analyses are done based on the pooled samples - i.e., comparing all households surveyed at baseline with all surveyed at midline. Sample weights were applied to adjust for the cluster sampling design and, in Dar es Salaam, for stratification. The sampling weight is a number that indicates the number of households in the city overall that each household in the sample represents. Applying sampling weights allows the team to adjust for that and tabulate descriptive statistics that are representative of the entire city, not reflective of the sampling design. Sample weights are equivalent to the inverse probability of being selected into the sample. The probability of a household's selection into the study is 8 divided by the number of households in the cluster, multiplied by 313 over the total number of clusters in the inventory for the stratum (or in the case of Morogoro, the number of clusters in the city since no stratification was done). In other words, it is the product of the probability of a cluster's selection and the probability of a household's selection. The sampling weight is 1/the probability of selection.

Questionnaires

Overview

Household Survey

Data Collection

Data Collection Dates

Start	End	Cycle
2013-05-01	2013-08-30	Baseline
2015-08-26	2015-11-10	Midline

Data Collection Notes

Data collection instruments were intensively pre-tested, refined, and piloted in the field. Field staff trainings culminated in a full-scale pilot of the survey protocols, household tracking, interviews using the final household survey tools, and collection of water quality samples from eligible households. A pilot survey was conducted in Kawe ward in August 2015 and all 36 potential field interviewers participated in the pilot testing. Research clearance was obtained for all activities through Social Impact's in-house Institutional Review Board (IRB) and additional in-country clearance was granted by the Tanzania Commission for Science and Technology (COSTECH), prior to the initiation of any field activities.
Field team supervisors were responsible for checking completed interview forms on field workers' tablets on a nightly basis after all interviews were complete for the day. Any questions were brought to the interviewer, after which responses were reconciled, and finalized interviews were then submitted to the server. In addition, 107 interviews were directly observed by a supervisor (55 in Morogoro and 52 in Dar es Salaam). A sample of 256 interviews were also back-checked (114 for Dar es Salaam and 142 for Morogoro). Following the completion of the interview, a supervisor would call the household to verify that they had indeed been interviewed, and were asked a few questions to verify the accuracy of responses recorded by the field worker during the initial interview. Supervisors also conducted 82 random spot checks (47 in Dar es Salaam and 35 in Morogoro) of surveys to monitor interviewers' performance and consistency of responses within the interview prior to submission to the server.
Overall, three quarters of respondents consented to audio audits (77% in Morogoro and 72% in Dar es Salaam). SI's data collection partner in Tanzania, CSR Group Africa, assigned a staff member to listen to a sample of the previous day's audio audits each day. This was used first to verify that interviewers were conducting the interview, and secondly as a teaching tool - for example, listening to the audio recordings prompted the local data collection partner to identify some interviewers that needed additional assistance in building rapport with respondents and shared such lessons as they arose with the entire field team during daily full-team briefings each morning.
A cognitive interviewing activity was conducted in early 2015, prior to the household survey, to test and refine survey-based measures of water supply. A set of 48 cognitive interviews was conducted in Morogoro and Dar es Salaam, 24 in each city, prior to the administration of the midline household surveys

Questionnaires

Household Survey

Data Collectors

Name	Abbreviation	Affiliation
CSR Group Africa, Limited		

Supervision

Key field staff for the midline data collection included a project team leader, field manager, data manager, and CATI supervisor. Field teams were composed of two Field Coordinators, supervisors, and enumerators. The training was attended by 36 qualified enumerators, and the 30 most successful candidates were hired. The Field Manager, who was responsible for overseeing all data collection teams in the field, was assisted by two Field Coordinators, who directly managed the survey and listing teams and collected the system-level data. This management team was based in Dar es Salaam and Morogoro for the entirety of data collection. During the data collection period, they oversaw the field teams (supervisors, enumerators, interviewers, laboratory staff) and coordinated all logistics, communicating any issues to higher management as needed. Water quality staff included two analysts hired by CSR Group Africa and trained by SI. Two field laboratories each in Dar es Salaam and Morogoro were setup, and SI provided the necessary laboratory equipment. Survey enumerators were trained in the water quality tesitng protocol. For the cognitive interviewing activity, the study team consisted of the SI evaluation team, a field coordinator, and two interviewers. The field coordinator was based in Dar es Salaam and traveled within Dar es Salaam and Morogoro to introduce the team to local officials prior to the cognitive interviews.

Data Processing

Data Editing

Data cleaning was mainly conducted by Social Impact with CSR Group Africa providing support and responses in relation to verifying different data quality assurance queries that were raised by Social Impact. The SurveyCTO software used for electronic data collection allows for real-time access to completed survey forms that have been submitted by field workers after the interviews are completed. Using these interim datasets, SI conducted regular and periodic checks of the data while data collection was ongoing. SI conducted regular checks to monitor progress and data quality during data collection and also conducted several rounds of more comprehensive data quality checks during and after data collection, including household survey data from SurveyCTO as well as water quality and logger data from mWater. Several rounds of data checks were communicated to local data collection partners, CSR Group Africa, and any inconsistencies were reconciled prior to analysis. Any changes to the raw datasets made through these processes have been documented in the files used to clean, manage, and analyze the data. Most issues raised through these exhaustive and iterative data quality assurance checks were relatively minor; any issues have since been reconciled, and are not expected to affect the data analysis in any meaningful way.

Other Processing

Data quality control for the midline survey involved several elements. First, constraints and validations were programmed directly into the electronic survey using SurveyCTO software. This enabled the team to restrict values into valid or logical ranges, control survey branching and logic, and activate relevant modules or lists of response options based on answers to previous questions. In addition, a number of warnings were programmed into the survey. These warnings are programmed to pop up on the interviewer's tablet screen based on entries to at least one question, and tell the interviewer that an entry for a certain question may be inconsistent or illogical, and asks the interviewer to go back and double-check the entry. Warnings serve as “soft” controls and can be useful in cases where valid responses cannot necessarily be anticipated a priori, or where it would not be preferable to strictly prohibit interviewers from entering certain values into a field, but where certain responses or ranges could be flagged for follow-up.
Household location and household member information was pre-loaded into the SurveyCTO software prior to household survey administration. Pre-loading the location information allowed for filtering when selecting a household's geographic location (city, district, ward, mtaa, and cluster) and corresponding ID (eight within each cluster), to eliminate the possibility of typing errors in data entry for location information. Pre-loading the baseline household member roster information also enabled SI to program an efficient set of verification modules; if a household being interviewed had also participated in the baseline data collection, interviewers were able to pull up the demographic information for each household member that was present at baseline and verify whether or not the individual was still a member of the household, and whether or not their information was still correct (e.g. whether educational attainment or marital status had changed). Pre-loading eliminates the lengthy process of filling in an entirely new household roster for all households surveyed that had also participated in the baseline and also greatly reduces the probability of data entry errors. Any new household members could also be added in as needed.

Data Appraisal

Estimates of Sampling Error

Sampling errors were calculated for all estimated quantities of indicators when tabulated to be representative of the city's population (i.e. when estimates are provided after sampling weights have been applied). The team calculated errors using the -svy- suite of commands in Stata 13. This module uses the Taylor linearization method of variance estimation for survey estimates that are means or proportions. The baseline and midline report contains a full account of the indicators estimated and the associated sampling errors. The baseline and midline reporst also contains a comprehensive account of all efforts taken by the evaluation team to avoid non-sampling errors (in measurement, sample frame coverage, and others).