

Tanzania - Water Sector 2013

Social Impact, Inc.

Report generated on: June 9, 2015

Visit our data catalog at: <http://microdata.worldbank.org>

Sampling

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 information about water supply by ward, clusters were chosen by stratified random sampling, 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. For each of the selected clusters, listing teams worked with local community representatives to enumerate all households in each EA and generate a complete sample frame of households. From each cluster's household list, 8 households per cluster (EA) were randomly selected for the household survey using 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 were selected for water quality testing (up to two per cluster). 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.

Deviations from Sample Design

If the listing team encountered any EAs in either city that had been demarcated strictly as an institution (e.g. hospital, school, jail) with only staff residing in the cluster, but had not been previously excluded from the sample frame, that EA was replaced by the next eligible EA from the list based on its random number, and the institutional cluster was excluded from the sample frame altogether. If community members or local officials declined to be involved in the surveying for any reason, that cluster was replaced. No deviations were made in the sampling procedures for the household survey. For the water quality testing, a much smaller sample of household taps was available for testing compared to initial expectations, so the eligibility for water quality tests was expanded at the beginning of these exercises to include shared sources in the community. Qualitative sampling was purposive and therefore was tailored to the specific objective of interviewing each type of respondent; while focus groups were initially planned to be a mix of males and females, after the first focus group the team decided to limit the participants to females only.

Response Rate

Response rates for the household survey in Dar es Salaam were above 87%, and above 92% in Morogoro. Overall, for the phone follow-up survey, response rates were 85% (round 1), 88% (round 2), and 90% (round 3); 81% of households overall participated in all 3 rounds while 90% participated in at least one round. Water quality test samples were drawn from household taps when available, or otherwise from other shared-source locations in the survey cluster (water quality results are intended to be representative at the cluster level). In Dar es Salaam, 95% of sampled clusters were covered by water quality tests, along with 99% of sampled clusters in Morogoro. Sampling for qualitative data collection components was purposive, in order to include specific types of respondents and target areas of each city with specific characteristics; this purposive sampling made extensive use of preliminary quantitative survey data and geospatial data. Qualitative research included, in total, 14 focus group discussions, 52 semi-structured interviews, and 10 key informant interviews.

Weighting

Sample weights were applied to the baseline dataset 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 the sampling design and tabulate descriptive statistics that are representative of the entire city, as opposed to just the households included in the survey. 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 the product of two ratios. The first ratio is 8 divided by the number of households in the cluster. The second ratio is 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
Phone Survey

Data Collection

Data Collection Dates

Start	End	Cycle
2013-05-01	2013-08-31	Baseline

DATA COLLECTION NOTES

Data collection instruments were pre-tested, refined, and piloted in the field. Field staff trainings culminated in a full-scale pilot of the survey protocols, including household listing exercises, random selection of households, interviews using the final household survey tools, and collection of water quality samples from eligible households. The full-scale pilot of the household survey took place in May 2013. Qualitative interview guides were piloted in July 2013 in Dar es Salaam, prior to administration in August 2013. SI participated in survey pre-testing, enumerator training, and piloting of all primary data collection protocols with the exception of the phone follow-up surveys. 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.

Data Collectors

Name	Abbreviation	Affiliation
Economic Development Initiatives, Ltd.	EDI	

SUPERVISION

Key field staff for data collection included a project manager, data manager/team leader, and field manager. In total, the field team was composed of two Field Coordinators, 12 listing supervisors, and 72 interviewers; field teams were split between the two cities. 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. The Data Manager made several visits to the data collection sites throughout the preparation, piloting, and data collection implementation phases to oversee activities and address any outstanding issues. Listing teams worked closely with community representatives to develop the sample frame. Water quality staff included two laboratory supervisors and six laboratory technicians, split between the two cities. A data processing coordinator and quality control coordinator supported the survey teams. Qualitative interviews were conducted by two facilitators. Training was required for all field staff, including supervisors and HQ staff, which included classroom training followed by a full-scale pilot of listing, sampling, and interviewing. For qualitative interviews, two experienced qualitative research consultants were hired by EDI. During the survey, supervisors were responsible for reviewing the interviews conducted by their teams at the end of each day. This included reviewing each completed interview manually, as well as running the automated validation checks to ensure that all issues picked up by the Surveybe validation rules have been identified. Any errors or inconsistencies detected at this stage were to be addressed with interviewers and reconciled before being uploaded into the database. EDI's policy was to conduct a full re-interview if the supervisors detected substantial discrepancies. In addition, supervisors were to complete direct observations of interviews several times per week, and upload a weekly report summarizing the performance of the interviewers and any actions taken in response. Supervisors were also required to conduct re-visit spot-checks in three to six households per week to ensure data collection quality using a standardized set of thirteen questions selected from the household questionnaire. This process was used to monitor interviewers and to provide constructive guidance as needed. Supervisors were also responsible for conducting re-interviews of the entire questionnaire if any interviewers were suspected of fraudulent behavior; no re-interviewers were necessary over the course of data collection.

Data Processing

Data Editing

EDI employed a data processing and quality control team, which was tasked with ensuring the quality of data collected through the Surveybe system. Daily checks of questionnaire data were conducted in Stata using a continually updated checking do-file, which flagged discrepancies and data inconsistencies. Each supervisor was provided a set of data checks to address with each team on a continuous basis. Data processing and quality control staff were primarily based at EDI headquarters, but were present in the field for several weeks during the beginning stages of each phase of data collection. This presence allowed them to participate in feedback sessions with interviewers demonstrating how data checks were conducted and how errors would be communicated to supervisors. The data processing team updated their checks periodically to accommodate new checks arising during the survey period, often in coordination with SI. SI's data quality monitoring strategy included continuous technical support to EDI over the entire period of data collection, field presence at all critical junctures during preparations for data collection, and several rounds of independent data verification with interim datasets provided throughout the survey period by EDI. SI wrote do files (using Stata 12) to monitor the quality of the data as it was received, updating them on a continuous basis and making adjustments after continuous communications with EDI's data manager and data processing teams, and ran through each of the datasets at numerous points between May and September 2013, communicating concerns or questions to EDI through a standard form used throughout the period of data collection. After the conclusion of the data collection, SI conducted a comprehensive data quality review of all datasets, inclusive of quantitative and qualitative datasets. SI submitted requests resulting from this review to EDI. EDI responded to these requests and subsequently delivered final datasets to SI via MCA-T. EDI produced several briefs on data quality assurance during data collection, and SI produced a data quality report for internal review by MCC.

Other Processing

EDI used its in-house electronic data collection system, called Surveybe, to administer the household survey data collection. Validation checks programmed into Surveybe display red flags on the screen to alert the enumerator when responses need attention. In this data collection effort, a total of 79 validation rules were programmed within the full baseline household survey and 57 in the phone survey. Interviewers were required to run the automatic validation checks once per screen, and again at the end of the interview. The interview would not be considered complete and would not be uploaded to the database if validation flags had not been addressed by interviewers. Interviewers also have the ability to input comments that provide explanation for seemingly inconsistent or outlier responses. All data from the Surveybe system were delivered to SI via MCA-T in Stata format. Final instruments were produced as PDF documents, exported from Surveybe, and included information on variable names, labels, response choices, validation rules, skip patterns, and the data table to which that data would export. Water quality test results were recorded on paper and subsequently transferred into a database uploaded through the Surveybe system. Photos of each water quality test slide read, with its own unique ID, were taken and provided to MCA-T and SI for QA purposes. System-level water quality tests were recorded directly into Microsoft Excel. Qualitative interviews were audio-recorded and transcribed directly into English in Microsoft Word; audio recordings were retained for QA purposes. Geospatial data collection by SI was transferred from hard copy to softcopy by SI's in-country GIS consultant and the scanned hard-copy documents were retained for QA purposes.

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 12. This module uses the Taylor linearization method of variance estimation for survey estimates that are means or proportions. The baseline report contains a full account of the indicators estimated and the associated sampling errors. The report 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).

Related Materials

Questionnaires

Household Survey Questionnaire Report

Title Household Survey Questionnaire Report
 Author(s) Social Impact
 Country Tanzania
 Language English
 Filename Household Survey Questionnaire Report - English.pdf

Phone Questionnaire Report

Title Phone Questionnaire Report
 Country Tanzania
 Language English
 Filename MCA_Phone_ENG_2006QxReport.pdf

Reports

Evaluation Design Report

Title Evaluation Design Report
 Author(s) Social Impact
 Country Tanzania
 Language English
 Filename design-tza-water-ie-jun12.pdf

Baseline Report

Title Baseline Report
 Author(s) Social Impact
 Date 2015-02-19
 Country Tanzania
 Language English
 Filename SI_MCC TZ Water IE_Baseline_Report_PUBLIC_FEB-19-2015.pdf

Technical documents

Baseline Data Collection Memo

Title Baseline Data Collection Memo
 Author(s) Social Impact
 Country Tanzania
 Language English
 Filename memobaselinedata-tza-water-final-jun13.pdf

Baseline Report Presentation

Title Baseline Report Presentation
Author(s) Social Impact
Date 2014-08-01
Country Tanzania
Language English
Filename SI_MCC TZ Water IE_Baseline_Report_Presentation_Aug2014.pdf

Baseline Read Me File

Title Baseline Read Me File
Country Tanzania
Language English
Filename TZA_Water_Baseline_README.pdf
