
DATE: 04/06/2020

TO: Algerlynn Gill, Millennium Challenge Corporation, Monitoring & Evaluation

FROM: Danae Roumis, Social Impact

RE: Transparency Statement of Evaluation Data and Results for Lesotho Metolong Program and Urban and Peri-Urban Water Activity

As per MCC guidance on responsible data management¹ and following Social Impact's requirements for responsible data management, the primary quantitative data associated with this evaluation is available for public use. Qualitative data and secondary quantitative and geospatial data associated with this evaluation will not be made available for public or restricted use.

The qualitative data associated with this evaluation unavailable for public or restricted use includes notes from:

- 1.) Key informant interviews with water sector stakeholders conducted as part of the process evaluation;
- 2.) Key informant interviews with chiefs, textile industry stakeholders, housing sector stakeholders, and small and medium enterprise sector stakeholders conducted as part of the summative evaluation; and
- 3.) Focus group discussions with households conducted as part of the summative evaluation.

Given "unknowns regarding de-identification and usability of qualitative data," MCC currently does not expect qualitative data to be prepared for public use.

The secondary quantitative and geospatial data furnished to the evaluation team associated with this evaluation includes:

- 1.) From the Water and Sewerage Company (WASCO):
 - a. Customer information, consumption data, and billing data from the EDAMS customer database from 2008 to March 2018
 - b. Geospatial data for urban water networks in Lesotho
 - c. Geospatial data for "as-built" location of works from the Lesotho Compact
- 2.) From the Lesotho National Development Corporation (LNDC):
 - a. Industrial employment data from January 2000 to September 2017
- 3.) From the Bureau of Statistics (BOS):
 - a. Geospatial data for boundaries of 2016 Census enumeration areas
 - b. Population information from the 2016 Census

In all of these cases, this data is the sole property of the institution who furnished it to the evaluation team. SI received the data as part of the evaluation with support from MCC without a signed agreement permitting public or restricted use by external parties.

¹ Available online here - <https://data.mcc.gov/evaluations/index.php/catalog/155>

The primary quantitative data from the evaluation's household survey questionnaire facilitates meaningful reproducibility of analysis in the Final Evaluation Report, with minor discrepancies in point estimates yielded by de-identification of the data that do not change the direction, significance, or interpretation of evaluation results. All necessary variables and variable transformations are available in this dataset to replicate evaluation results. In almost all cases, raw versions of variables are also provided in the dataset and the corresponding variable generation code submitted as part of the data package can be used to reproduce the construction of analysis variables and covariates or test the sensitivity of findings to alternative specifications.

The only exceptions to this rule are below:

- 1.) In many cases, there were only a few connections per month per town in the WASCO EDAMS database, such that someone with access to this database could reidentify a household based on the combination of its town and installation date alone. Thus, we have dropped installation dates from the dataset. These dates were not used in any evaluation analysis other than the construction of eligibility variables for the evaluation. However, users will not be able to use these variables for other research purposes or to reconstruct eligibility due to their absence.
- 2.) Sampling weights for the customer survey were derived from customer information from the WASCO EDAMS database. As we are unable to publish data from this source, it will not be possible for users to reconstruct these sampling weights. The weights themselves are included in the public use file to replicate evaluation results.
- 3.) While the child-level data collected as part of the illness module in the household survey are available, there was an additional "per-caretaker-per-child" sub-module (i.e. a repeat-group within a repeat-group) that inquired as to the circumstances of each household member who cared for a specific sick child over the past two weeks that is not available. Only two households qualified to complete this module, and corresponding information is collapsed to the child-level in the illness supplemental dataset. This collapse does not prevent the user from replicating any evaluation analysis. It was ultimately not used for any analysis.
- 4.) The population and geospatial data furnished by WASCO, LMDA, and the BOS was used to determine eligibility for the study and to calculate variables used in the final treatment and outcome models. Namely, it was used to calculate the distance between each household and the local water network and the proportion of households in each enumeration area that were WASCO customers prior to interventions. These variables are available in their final form but cannot be reconstructed without the secondary data inputs from these sources.

For more detail on the proposed data preparation process, please reference the Data De-Identification Worksheet in this data package. For reference, the "1_de_id_analysis_MASTER CONTROL" do file is inclusive of all other do files submitted and re-runs analysis from start to finish using the other do files.