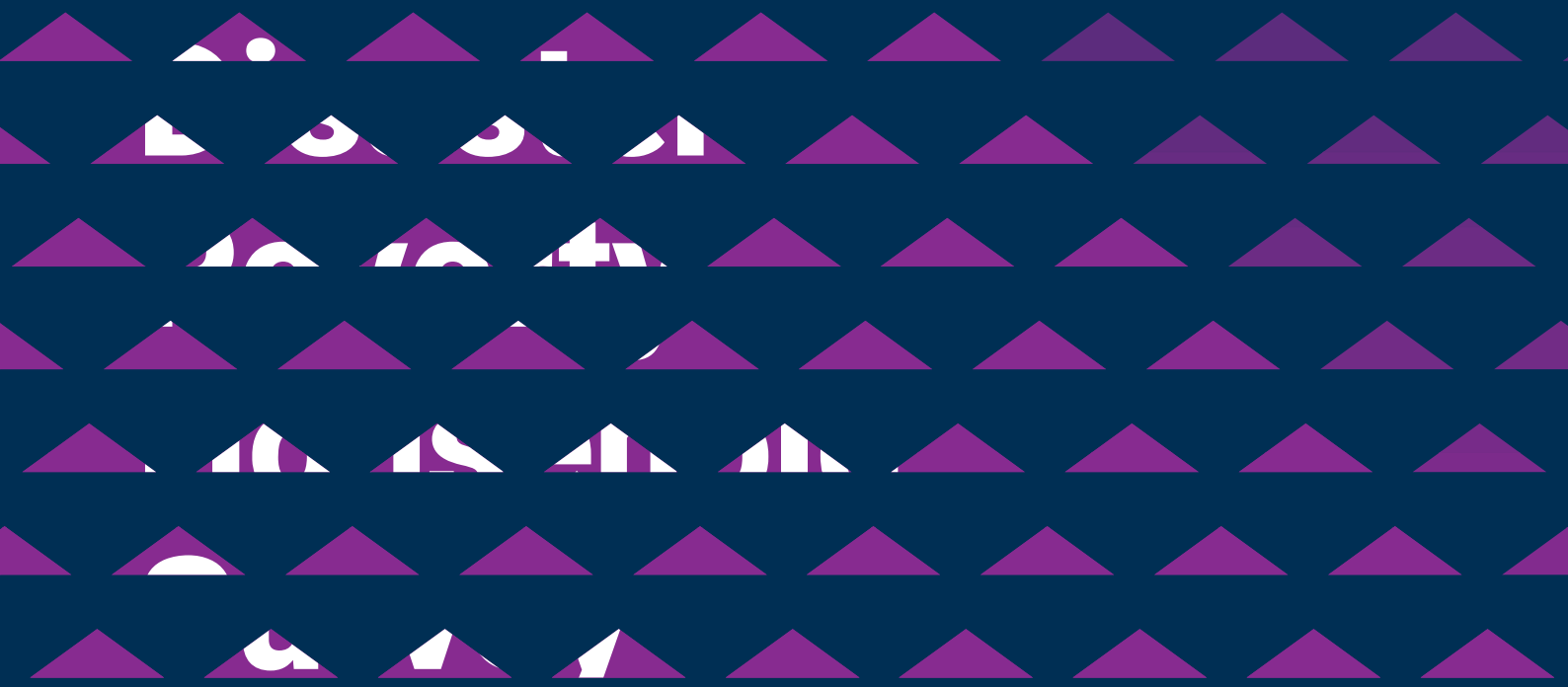
TECHNICAL NOTE

Dar es Salaam, Tanzania



GFDRR
Reduction and Recovery

1 Overview

Content of this document: This document provides information about the Disaster Poverty Household Survey (DPHS). It describes the DPHS series, the survey and sampling design and the questionnaire used, and it discusses some data considerations, including outlier treatment and anonymization process.

Objective of the survey: The DPHS is designed to collect information that can be used to assess the relationship between disasters (exposure, vulnerability, and capacity to recover) and poverty in the urban environment. The data can be used to explore policy-relevant research topics related to climate change adaptation, urbanization, urban poverty, and more.

Content of the data: DPHS data contains information on household characteristics, household expenditure, living conditions and household experience with disasters. Household characteristics include household size and member level information on religion, education and labor. Household expenditure is collected using the Survey of Well-being via Instant and Frequent Tracking (SWIFT) methodology, which estimates household income (or consumption expenditure) based on non-monetary variables that are highly correlated with poverty. Information on living conditions covers housing quality, asset ownership, access to services and jobs, rent and housing costs and tenure arrangements. Information on experiences with disasters includes direct and indirect impacts of historic disasters on household assets, education, health and labor access, as well as impacts on public services. There is also information on coping behaviors and perception of risk of future exposure. The DPHS can be customized to collect information on different disasters. So far, it has mainly focused on the impacts of urban flooding.

Dar es Salaam Application: The DPHS in Dar es Salaam was conducted in two rounds in November-December 2017 and in September 2018, with the objective to assess the role of poverty and other social factors in urban flooding in the city. The survey data collected in 2017 focused on exposure to frequent flooding, while the follow up survey in 2018, targeting the same households, focused on the impact of a flood event that happened in April 2018. During the follow up survey in 2018, additional households were also added to the sample. The data collected is representative at the city level and overrepresented in areas that are flood prone.

This project was a collaborative effort between Global Facility for Disaster Reduction and Recovery (GFDRR), the Tanzanian Urban Resilience Program (TURP), the Poverty Global Practice and Urban, Disaster Risk Management, Resilience and Land Global Practice (GPURL). Data collection was carried out by UDA Consulting under World Bank supervision.

Data files and other resources

- DPHS_Dar_es_Salaam_Data_2017: DPHS data in STATA format (.dta)
- Additional_Households_Dar_es_Salaam_2018: DPHS data in STATA format
- Follow_up_Dar_es_Salaam_2018: Follow up survey in 2018 data in STATA format
- DPHS_Dar_es_Salaam_SWIFT: SWIFT (household expenditure) data in STATA format (all households)
- DPHS_Dar_es_Salaam_Questionnaire: DPHS Questionnaire in Excel
- Follow_up_Dar_es_Salaam_Questionnaire: Questionnaire for the follow up survey data in Excel

Citation requirements:

The World Bank. Disaster Poverty Household Survey (DPHS), Dar es Salaam, Tanzania 2017-2018. Dataset downloaded from microdata.worldbank.org on [date].

2 The survey

2.1 Description

Name of the study: Disaster Poverty Household Survey (DPHS), Dar es Salaam, Tanzania

Geographical coverage: Dar es Salaam

Sample size: 1335 (1053 in 2017 and 282 in 2018)

Date of the survey: November and December 2017 (DPHS) and September 2018 (follow up and additional household surveys)

Primary Investigators: Alvina Erman (World Bank), Silvia Malgioglio (World Bank), Nobuo Yoshida (World Bank), Stephane Hallegatte (World Bank)

Collaborators: TURP, GFDRR, Poverty Global Practice, GPURL and UDA Consulting

Funding: TURP and Global Facility for Disaster Reduction and Recovery

Related reports: Erman, et al (2019)

2.2 Sampling design

The selection of households in the survey design had two objectives. First, to select a sample that represents the population of Dar es Salaam and second, to interview enough people who had experienced floods to be able to detect patterns in their socio-economic characteristics.

The sample size was selected to confidently represent the population of Dar es Salaam given the income level and income distribution. Accordingly, a sample size of 105 EAs and 10 households per EA were selected using Probability Proportion to Size (PPS). In 2018, 28 EAs to the original sample as part of an additional round of data collection.

To capture enough households that had experienced floods, a flood risk stratum was designed using the Ramani Huria community flood map[1]. EAs were categorized according to three flood risk strata, i.e., “no risk”, “low to medium risk” and “high risk”, depending on how much of the EA was covered by the flood layer in the map. This categorization of the city was used to oversample in high risk and low-to-medium risk areas by selecting more of those EAs compared to the population living there. Finally, all the selected households were randomly drawn within each EA using satellite imagery.

Sampling weights were calculated to compensate for the oversampling in high-risk areas. When applying the sample weights, the dataset is representative at the city level.

2.3 Survey implementation

The DPHS was conducted in two waves. The main survey was conducted in November and December 2017 using the original sample of 1058 selected households. Shortly after in April 2018, there was a significant flood event affecting Dar es Salaam. To evaluate the impacts of the flooding, a phone-based follow up survey was carried out in September 2018. Out of the 1058 households, 419 were reached and interviewed[2]. In parallel, the original sample size was increased with 282 additional households that were sampled from high-risk and low-to-high risk areas to further increase precision of estimates regarding this population group. Interviews with additional households were carried out in-person.

[1] The community maps were created between 2015-2017 and used information from community members to identify flood prone areas. https://ramanihuria.org/en/
[2] Among the 730 households for which phone numbers had been obtained in 2017, 419 were reached and agreed to the follow up interview, which corresponds to a response rate of around 60%.

3 Questionnaire modules (main questionnaire)

- K: Household information: pre-filled before the interview
- B: Household member roster
    - Educational attainment
    - Labor participation
- M: Asset ownership
- H: Housing and services
    - Tenure arrangements
    - Housing costs and rent
    - Tenure security
    - Housing quality
    - Access to services
    - Remittances
- J: Weather information
- W: Preventive measures flood
- E: General information on experience with flooding
- F: Information on specific floods
- U: Perception of risk and community engagement
- C: Household enterprise
- I: Investments in housing
- L: Food Insecurity (Reduced Coping Strategy Index CPI-R)

4 Data considerations

4.1 Anonymization of the dataset

Protecting the privacy of survey respondents is of the outmost importance to the World Bank. To make sure the data cannot be used to identify individual households in the dataset, a technique of statistical disclosure control (SDC), as described in Benschop et al. (2021), was applied. It helped identify variables that included unique information about households. After identifying the high-risk variables, necessary adjustments were made to make sure the SDC analysis provided satisfactory results, i.e., low risk of re-identification. Results can be shared upon request. The following data editing was done for anonymization purpose:

- Precise location data, such as GPS coordinates, were dropped
- Personal information, such as name, citizenship and phone number were dropped

- Information on from which region or country the respondent moved from before settling in current dwelling and where respondent was born was categorized into “in Dar es Salaam” and “outside Dar es Salaam” to protect privacy while preserving valuable data. District level information on origin was dropped.
- Household size exceeding seven household members was categorized as “above 7 members”
- Household member information for 7th member and above was dropped to avoid reconstruction of the household size variable.

4.2 Outlier treatment

Continuous variables may present some measurement errors. A technique of outlier's treatment is recommended. Some of these variables are:

- *h5:* How much does this household pay to rent this residence per month? (in *TSh*)
- *h8a:* If a friend wanted to rent a property like this in the same neighborhood, how much would he/she have to pay per month? (in *TSh*)

An established method to identify outliers is to tag the observations that deviate from the mean by a set number of standard deviations. Three standard deviations are commonly used. Figure 1 includes STATA code that can be used to tag outliers[3].

Figure 1: Codes for the identification of outliers for the variable *h5 (rent)*

[3] Additional checks may be conducted to analyze the presence of outliers. The technique in Figure 1 assumes that the distribution of the variable is normal. This may not be the case, even after using a logarithmic transformation. Other transformations and for which kinds of variables to use them are explained in Ravallion (2017). Outliers may influence the mean and the median of the distribution. More robust methods of outlier treatment may be necessary, for instance, the median absolute deviations (MAD) method (Belotti et al., 2021, Rousseeuwand Croux, 1993).

```
* Create dummy for outliers
foreach var of varlist h5 {
    quietly summarize `var'
    g Z_`var'=(`var' > 3*r(sd)) if `var'<.
    list `var' Z_`var' if Z_`var'==1

}
/*        +---------------+
          |     h5   Z_h5 |
          |---------------|
     227. | 350000      1 |
     746. | 400000      1 |
     764. | 500000      1 |
     983. | 300000      1 |
          +---------------+

*/
```

The outliers can then be replaced in function of the variable studied. For instance, for the variable *h5 (rent),* a hedonic regression can be applied to identify drivers of rent, which can be used to produce predicted rent values that can then replace the outliers identified.

5 SWIFT Methodology

Household consumption data are costly to collect and significantly increases the duration of interviews. Beyond budgetary and data processing issues, it also reduces the quality of the data by reducing the space available for the other questions in the survey and by increasing the risk of survey fatigue of respondents. To avoid these issues, the survey adopted the SWIFT approach to estimate household expenditures and poverty rates. The SWIFT methodology collects household data using a short list of questions that can be integrated into the questionnaire and computes an estimated household income (or consumption expenditure) based on non-monetary variables that are highly correlated with poverty. SWIFT uses survey-to-survey imputation based on official household data and produces estimates comparable to official data. More details on SWIFT are provided in Yoshida et al. (2021).

The SWIFT variables are provided in the dataset *DPHS_Dar_es_Salaam_SWIFT*. It can be matched with the *DPHS_Dar_es_Salaam_Data_2017* and *Additional_Households_Dar_es_Salaam_2018* using the variable *hhid*.

References

BELOTTI, F., G., VECCHI, G., AND MANCINI (2021): “OUTDETECT: Stata module to perform outlier detection and diagnostics for welfare analysts,” *Statistical Software Components* S458932, Boston College Department of Economics.

BENSCHOP, T. AND WELCH, M. (n.d.): “Statistical Disclosure Control for Microdata: A Practice Guide”, Retrieved 30 March, 2022, from https://sdcpractice.readthedocs.io/en/latest/

ERMAN, A. E., TARIVERDI, M., OBOLENSKY, M. A. B., CHEN, X., VINCENT, R. C., MALGIOGLIO, S., RENTSCHLER, J., HALLEGATTE, S., & YOSHIDA, N. (2019). Wading Out the Storm : The Role of Poverty in Exposure, Vulnerability and Resilience to Floods in Dar es Salaam. *Policy Research Working Paper.* World Bank, Washington, DC.

RAVALLION, M. (2017): “A concave log-like transformation allowing non-positive values,” *Economics Letters*, 161, 130-132.

ROUSSEEUW, P. J., CROUX, C. (1993): “Alternatives to the median absolute deviation,” *Journal of the American Statistical association*, 88(424), 1273-1283.

YOSHIDA, N., X. CHEN, S. TAKAMATSU, K. YOSHIMURA, S. MALGIOGLIO, S. SHIVAKUMARAN, K. ZHANG, D. ARON. (2021): “The Concept and Empirical Evidence of SWIFT Methodology,” The World Bank.