TECHNICAL NOTE

Accra, Ghana

Disaster Poverty Household Survey

WORLD BANK GROUP

GFDRR
Global Facility for Disaster Reduction and Recovery

1 Overview

Content of this document: This document provides information about the Disaster Poverty Household Survey (DPHS) in Accra, Ghana. 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 to assess the relationship between disaster risk (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.

Accra application: The DPHS in Accra, Ghana was collected in May and June 2017 in slum areas across nine neighborhoods in the city. The survey focused on the impacts of a major flood event that happened in June 2015 in Accra and how the impacts related to the poverty status of households, focusing on exposure, vulnerability and capacity to recover.

This project was a collaborative effort between Global Facility for Disaster Reduction and Recovery (GFDRR), the Poverty Global Practice and Urban, Disaster Risk Management, Resilience and Land Global Practice (GPURL). The Institute of Statistical, Social and Economic Research (ISSER) of the University of Accra carried out the data collection under World Bank supervision.

Data files and other resources:

- DPHS_AccraGhana_Data_2017: DPHS data in STATA format (.dta)

- DPHS_AccraGhana_SWIFT: SWIFT (household expenditure) data in STATA format
- DPHS_AccraGhana_Questionnaire: DPHS Questionnaire in excel
- DPHS_AccraGhana_Neighbourhoods: A folder with the shapefile of neighborhoods

Citation requirements:

The World Bank. Disaster Poverty Household Survey (DPHS), Accra, Ghana 2017. Ref: DPHS_AccraGhana_Data_2017. Dataset downloaded from microdata.worldbank.org on [date].

2 Survey design

2.1 Description

Name of the study: Disaster Poverty Household Survey, Accra, Ghana

Geographical coverage: 9 slums areas in Accra, Ghana

Sample size: 1,006 households

Date of the survey: May-June, 2017

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

Collaborators: GFDRR, Poverty Global Practice, GPURL and University of Accra

Funding: GFDRR

Related reports: Erman et al. (2018, 2020)

2.2 Sampling design

The sample selection stratifies the targeted slums by flood proneness and the level of poverty (Erman et al., 2018) as the following:

1. Slum areas were identified by combining the definition for informal settlement used by Accra Metropolitan Assembly (AMA) and UN Habitat (2011) and a slum index score developed by Engstrom et al. (2017). Enumeration areas (EAs) were added to the sample frame if they were defined as being in a slum area using the following definition: i) they were fully inside the areas defined as informal settlement according to AMA and UN Habitat's definition and ii) had a slum index value higher than 0.7.

2. Enumeration areas in the sample frame were categorized as low poverty and high poverty by using a neighborhood-level poverty estimate created by Engstrom et al. (2017).

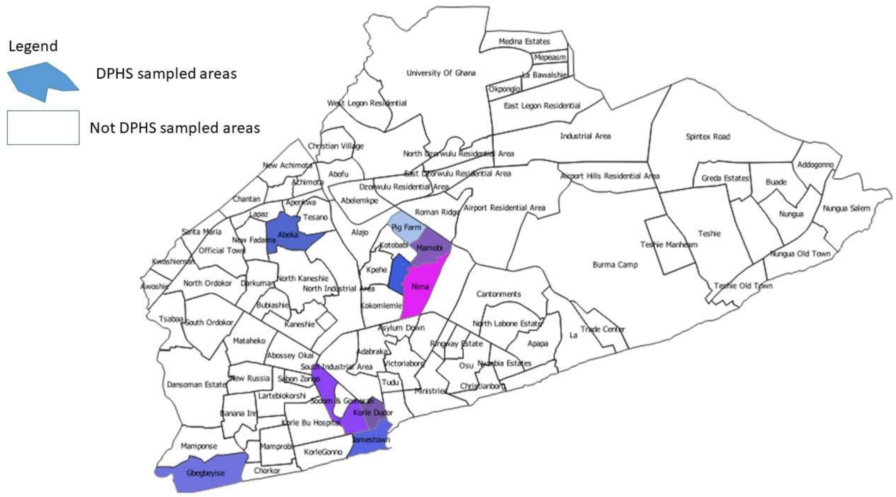
3. Enumeration areas in the sample frame were also categorized as flood-prone and not flood-prone using average elevation levels in the enumeration area. High flood risk areas are defined as below 17.5 meters (based on average elevation of areas flooded in the 2015 flood) and low risk areas as above 35 meters (the elevation level, above which there were no reported flooding during the 2015 flood).

4. Four neighborhoods in which all EAs were considered high risk and 4 neighborhoods in which all EAs were considered low-risk and one neighborhood with a mix of high and low-risk EAs were selected for the sample frame. In all selected neighborhoods, all EAs were defined as slum areas. The neighborhoods selected were Korle Lagoon Area, Jamestown, Gbegbeyise and Korle Dudor as high flood risk areas, and Abeka, Accra New Town, Mamobi, and Nima as low flood risk areas and Pig Farm, which includes both high and low flood risk areas. Neighborhoods are indicated in Figure 1 in a map of Accra. This administrative division was extracted from Engstrom et al. (2013).

5. The EAs in the selected neighborhoods were stratified into four categories: i) high flood risk and high poverty incidence; ii) low flood risk and high poverty incidence; iii) high flood risk and low poverty incidence; iv) low flood risk and low poverty incidence, of all selected neighborhoods.

6. Two-stage sampling was applied; 12 EAs per strata were selected using Probability Proportion to Size (PPS) and then 20 households per selected EA were selected using random sampling after listing. The sample size was determined using power calculations.

Figure 1: Map of neighborhoods in Disaster Poverty Household Survey in Accra, Ghana



The shapefile of the Accra neighborhoods can be found in the folder DPHS_AccraGhana_Neighbourhoods, among the resources made available. The neighborhood shapefile can be matched with the surveyed neighborhoods in the DPHS dataset (DPHS_AccraGhana_Data) through the key variable *neighbourhood_code*.

2.3 Representativeness of the sample

The objective of the sampling strategy was to be able to find systematic differences in characteristics and behaviors by flood risk exposure and poverty incidence among population living in slum areas. The data can be used to explore the relationship between poverty and flood risk for slum dwellers in Accra. The data is not representative at the city level, or for population living in non-slum areas. Generalizations based on the data made for population groups that do not fit into the sample characteristics should be made with great cation.

3 Questionnaire

- K: Household information: pre-filled before the interview
- B: Household member roster
  - Educational attainment
  - Labor participation
  - Social protection
  - Savings
  - Health
- M: Asset ownership
- H: Housing and services
  - Tenure arrangements

  - Housing costs and rent
  - Tenure security
  - Housing quality
  - Access to services
  - Agriculture
  - Remittances
- C: Household enterprises
- I: Investments in housing
- A: Questions regarding the June 2015 floods
  - Housing damages
  - Other assets
  - Labor impacts
  - Education impacts
  - Infrastructure impacts
  - Prices
  - Health impacts
  - Agriculture impacts
  - Income losses
  - Coping strategies
  - Assistance
  - Evacuation
- O: Early warning and indirect and long-term effects of 2015 flood
- W: Preventive measures
- V: Other floods
- D: Other natural shocks
- P: Idiosyncratic shocks
- U: Perception of risk
- L: Food Insecurity (Reduced Coping Strategy Index CPI-R)

4 Data considerations

4.1 Data anonymization

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
- Identifying information, such as name, birth date and phone number were dropped
- Furthermore, the number of reported religions was reduced from 8 to 3 categories, the number of ethnicities from 9 to 4 categories and 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 treatment is recommended. Some of these variables are:

- *b11d_commuting_time*: How long does it take to get to the workplace? (minutes)
- *h34_cash_send_3m*: Over the last 3 months, how much did you send cash or gifts to family
- *h6c_how_much_rent*: How much do you pay in rent each time? (in GH¢)

Figure 2: Codes for the identification of outliers for the variable *h6c_how_much_rent*

```
*Create a dummy for outliers
foreach var of varlist h6c_how_much_rent {
                                quietly summarize `var'
                                g Z_`var'= (`var' > 3*r(sd)) if `var' < .
                                list `var' Z_`var' if Z_`var' == 1
                            }

*Result:

/*
      +-------------------------------------------+
      | h6c_how_much_rent   Z_h6c_how_much_rent   |
      |-------------------------------------------|
 224. |    8640                                 1 |
 261. |   11520                                 1 |
 353. |   21600                                 1 |
      +-------------------------------------------+
*/
```

The outliers can then be replaced using regression analysis and prediction. For instance, for the variable *h6c_how_much_rent* (amount of 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.[1]

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_AccraGhana_SWIFT*. It can be matched with the *DPHS_AccraGhana_Data_2017* at the household level using the key variable *hhid*.

[1] Additional checks may be conducted to analyze the presence of outliers. The technique in Figure 2 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).

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/**

ENGSTROM, R., D., PAVELESKU, T., TANAKA, A., AND WAMBILE (2017): "Monetary and non-monetary poverty in urban slums in Accra: Combining geospatial data and machine learning to study urban poverty," Work in Progress, The World Bank.

ENGSTROM, R., OFIESH, C., RAIN, D., JEWELL, H., AND WEEKS, J. (2013): "Defining neighborhood boundaries for urban health research in developing countries: A case study of Accra, Ghana", *Journal of Maps*, 9(1), 36-42.

ERMAN, A., MOTTE, E., GOYAL, R., ASARE, A., TAKAMATSU, S., CHEN, X., MALGIOGLIO, S., SKINNER, A., YOSHIDA, N., AND HALLEGATTE, S. (2018): "The road to recovery: the role of poverty in the exposure, vulnerability and resilience to floods in Accra," Policy Research Working Paper; No. 8469. World Bank, Washington, DC.

——— (2020): "The Road to Recovery the Role of Poverty in the Exposure, Vulnerability and Resilience to Floods in Accra," *Economics of Disasters and Climate Change*, 4(1), 171–193.

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.