

TECHNICAL NOTE

---

# Cap-Haïtien, Haiti

# 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). 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 poverty based on household characteristics which are highly correlated with household wellbeing. 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.

**Cap-Haïtien application:** The DPHS in Cap-Haïtien was conducted in October and November 2018 in the metropolitan area of the city. The focus of the data collection was to capture information on exposure to flooding in Cap-Haïtien as well as household preparedness including access to early warning systems and the use of coping strategies.

The Poverty and Equity Global Practice designed, conducted, and managed the project in coordination with Global Practice of Urban, Resilience, and Land (GPURL). The Global Facility for Disaster Reduction and Recovery (GFDRR) financed the project. The Interuniversity Institute for Research and Development (INURED) carried out the data collection under World Bank supervision. The municipality of Cap-Haïtien also supported data collection.

### Data files and other resources

- DPHS\_CapHaitienHaiti\_Data\_2018: DPHS data in STATA database format (labels in English)
- DPHS\_CapHaitienHaiti\_SWIFT: SWIFT (household expenditure) data in STATA database format
- DPHS\_CapHaitienHaiti\_Questionnaire: Questionnaire for the DPHS data in excel (in French)

**Citation requirements:**

The World Bank. Disaster Poverty Household Survey (DPHS), Cap-Haïtien, Haiti 2018. Dataset downloaded from [microdata.worldbank.org](https://microdata.worldbank.org) on [date].

## 2 The survey

### 2.1 Description

**Name of the study:** Disaster Poverty Household Survey, Cap-Haïtien, Haiti

**Geographical coverage:** Metropolitan area of Cap-Haïtien

**Number of observations:** 588 households

**Date of the survey:** October and November 2018

**Primary Investigators:** Sering Touray (World Bank), Emilie Perge (World Bank)

**Collaborators:** GFDRR, Poverty Global Practice and GPURL from the World Bank, and INURED

**Funding:** GFDRR

**Related reports:** Touray and Perge (2020)

### 2.2 Sampling design

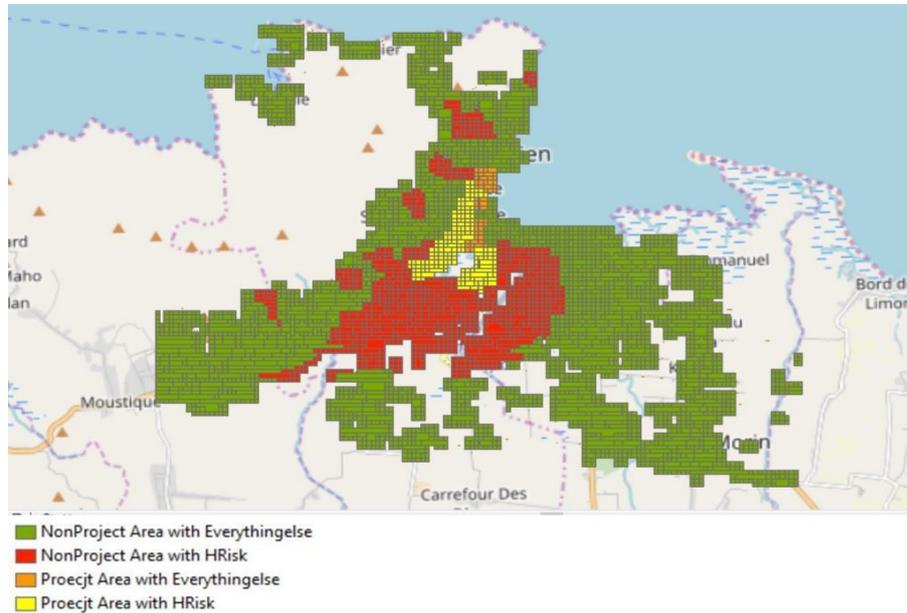
A stratified two-stage sampling strategy was applied to ensure the representativeness of the survey results. The sampling frame was constructed using the consumption aggregates from the 2012 *Enquête sur les Conditions de Vie des Ménages après le Séisme* (ECVMAS) collected by the *Institut Haïtien de Statistique et d'Informatique* (IHSI). The first stage identifies all applicable Primary Sampling Units (PSUs); and the second stage selects households within selected PSUs.

PSUs are categorized based on two spatial criteria: location the project area and location in areas with high flood risk. The project area in Cap-Haïtien refers to the areas benefiting from the Municipal Development and Urban Resilience project (MDUR, P155201). The high-risk areas are identified based on hazard maps and refer to areas with ‘moderate to high/strong’ and ‘strong to very strong’ risks of floods (Guillande, 2015). Based on these two criteria, four strata are constructed: project areas with high/moderate risk of floods; project areas with low/no risk of floods; non-project areas with high/moderate risk of floods; and non-project Areas with low/no risk of floods.

To calculate the probability of selection of households, WorldPop data is used to assign number of households per PSU. In the second stage, a listing of all households within a given PSU is conducted to determine the probability of household selection. By design, 120 PSUs were surveyed and 5 households per PSU were selected.

Weights are applied to ensure that the estimates obtained from the survey data are representative of the population of interest.

**Figure 1:** Sampling frame Cap-Haïtien



Source: Touray and Perge (2020, pp. 29)

### 3 Questionnaire modules

The questionnaire contains twenty modules with questions at the household and individual level. The questionnaire is only available in French<sup>1</sup>.

At the household level, questions include housing characteristics, tenure status, and asset ownership. At the individual level, there are questions on education, employment, and unemployment. Additionally, questions on the ownership, use, and coverage of mobile phones are asked to the household head.

Questions on flooding impact, recovery and coping are asked at the household level. They include questions about the experience with frequent floods and perception of risk to future flooding. Impacts of flooding on housing, assets, consumption, access to public services and transport, work, education, and health are asked. Finally, questions on coping strategy and awareness of emergency warning systems are also added.

## 4 Data considerations

### 4.1 Anonymization of the dataset

Protecting the privacy of survey respondents is of the utmost 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 and phone numbers were dropped
- Categories for type of dwelling were reduced from 7 to 4.

### 4.2 Outlier treatment

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

- k5: How long did it take to make reparations?
- k6: How much was the cost of reparations?

An established method to identify outliers is to tag the observations that deviate from the mean

---

<sup>1</sup> However, the variable and value labels in the dataset are in English.

by a set number of standard deviations. Three standard deviations are commonly used. Figure 2 includes STATA code that can be used to identify outliers<sup>22</sup>. Outliers can then be removed or replaced, using different methods.

**Figure 2:** Codes for the identification of outliers for the variable *k6* (repairs costs)

```

* Create dummy for outliers
  foreach var of varlist k6 {
    quietly summarize `var'
    g Z_`var'=(`var' > 3*r(sd)) if `var'<.
    list `var' Z_`var' if Z_`var'==1
  }
  /*
  +-----+
  |      k6      Z_k6 |
  +-----+-----+
  238. |    20000      1 |
  239. |    23000      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 data on expenditure are in this dataset: *DPHS\_CapHaitienHaiti\_SWIFT*. The *DPHS\_CapHaitienHaiti\_SWIFT* can be matched with the *DPHS\_CapHaitienHaiti\_Data\_2018* using the key variable *hhid*.

<sup>22</sup> 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

## 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/>

GUILLANDE, R., (2015), Caractérisation et cartographie du risque inondation et de submersion marine sur l’agglomération du Cap-Haïtien, rapport SIGNALERTI n° rep-CIAT – 04-15-001v3, April 2015 retrieved 05/16/2022 at <http://ciat.gouv.ht/sites/default/files/docs/Rep-FINAL-CIATLIDAR-04.15.001V3.pdf>

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

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

TOURAY,S.; PERGE,E. B., 2020, Poverty and Floods in Cap-Haïtien (English). Washington, D.C. : World Bank Group.

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. Mimeo.