Basic Information Document

São Tomé e Príncipe

COVID-19 Household Monitoring Survey

(COVID-19 HMS)

Version 4 (May 25, 2021)

Contents

1. Introduction

The purpose of this document is to provide detailed information on São Tomé e Príncipe (STP) COVID-19 Household Monitoring Survey (COVID-19 HMS), implemented by the National Statistical Institute (INE) from July 2020. To monitor how the COVID-19 pandemic affects STP's economy and population and to substantiate response policies with data, INE, with technical support from the World Bank (WB), has designed and conducted two rounds of a telephone Household Monitoring Survey (HMS). With support from the United Nations, the first round survey was expanded to include a questionnaire aimed at informal businesses.

The World Bank is providing support to countries to help mitigate the spread and impact of the new coronavirus disease (COVID-19). One area of support is for data collection to inform evidence-based policies that may help mitigate the effects of this disease. The STP COVID-19 HMS is a high-frequency phone survey that monitors the economic and social impacts of the COVID-19 pandemic on Santomean households and responses to it in terms of access to basic food, access to educational activities during school closures, employment dynamics, family income and livelihoods, loss of income, and food security.

The first cases of Covid-19 in São Tomé and Príncipe (STP) were registered on April 6, 2020, and as of the writing of this report, 2,338 cases and 37 deaths from the disease have been recorded. During this period, the STP government adopted measures to mitigate the risk of the virus spreading, such as the total or partial closure of schools, restaurants, bars, airspace, and commercial establishments in general. After June 2020, with the gradual relaxation of those measures, STP mobility returned to levels close to February of the same year. The first round of the survey takes place shortly after the reopening in July and August 2020. The second round takes place in January and February 2021, with the relaxation measures were already in place for six months. Commercial establishments were already allowed to operate at more flexible hours, and schools had already resumed classes.

2. Sample and Weights

2.1 Survey Sample

The STP COVID-19 HMS sample consists of a subsample of the Multiple Indicator Cluster Survey (MICS) carried out by INE in collaboration with UNICEF in 2019. Households with access to a telephone are represented in the HMS, covering urban and rural areas in all STP regions. The HMS called all households with a valid telephone number listed in MICS, completing 1,025 interviews (413 in rural areas and 612 in urban areas).

Among the 3,426 households interviewed in the MICS 2019, 1,400 (40.8%) provided at least one phone number. From these, 1,081 were successfully contacted by INE interviewers and 1,025 accepted and answered all the questions made in the first round of the HMS.

To mitigate bias in a sample that contains only households with a working telephone, a procedure for adjusting the sample weights was carried out using the Propensity Score Weighting (PSW) methodology. Following this procedure, the HMS results were brought closer to the national representativeness of surveys carried out in person, such as MICS 2019.

2.2 Survey Weights

Due to the way in which the HMS sample was constructed and the problems inherent in a survey carried out exclusively by telephone, it is expected that the statistical results obtained are potentially biased. This is because an inquiry like this does not reach households without a valid phone number, non-randomly excluding a portion of the population. The purpose of this subsection is to present the methodology used to mitigate the bias in the results obtained by the HMS.

2.2.1 Methodology

The methodology used to mitigate potential bias in the HMS sample was Propensity Score Weighting (PSW). This methodology assumes that the participation of a family in a survey carried out by phone depends on some observable characteristics of these families. This approach was originally developed to make a control group comparable to a treatment group (Rosembaum and Rubin 1983 and 1984), but has recently been applied to make statistics from a phone or web survey comparable to those of a nationally representative survey (for example , Terhanian, et al., 2000, Schonlau, et al., 2006, Lee, 2006 and Cappaci et al., 2018). The potential bias of inquiries carried out by phone or web comes from the fact that the profile of the interviewed families may be concentrated in certain groups of the population, distancing themselves from a sample that represents the entire country.

To conduct a PSW, it is necessary to find a survey of national representativeness that will serve as a model to be followed by the potentially biased survey. The objective of the PSW is to estimate new sample weights so that the new weighted statistics are similar to the nationally representative survey statistics.

2.2.2 Step by step of adjusting the STP HMS sample weights using PSW methodology

1. **Identifying a model survey:** the first step in conducting the PSW is to identify a nationally representative survey that will be used as a model. In the case of STP, the model survey used is MICS 2019. Table 1 shows the geographic distribution of the families interviewed at HMS and MICS 2019.

Table 1 - Geographic distribution of the HMS and MICS 2019 sample

Region	**HMS Covid-19**			**MICS 2019**		
	Rural	Urban	**Total**	Rural	Urban	**Total**
DISTRITO DE ÁGUA GRANDE	0	158	**158**	0	734	**734**
DISTRITO DE MÉ-ZÓCHI	104	50	**154**	479	264	**743**
REGIÃO NORTE OESTE	97	216	**313**	384	364	**748**
REGIÃO SUL ESTE	130	120	**250**	319	421	**740**
REGIÃO AUTÓNOMA DO PRÍNCIPE	82	68	**150**	302	159	**461**

Total	413	612	**1025**	1484	1942	**3426**

Source: HMS Covid-19 and MICS 2019

2. **Identifying variables in common:** it is necessary to choose variables present in both surveys that may be related to the participation of households in the interview. Such characteristics are generally related to demographic, consumption and household structure characteristics. Table 2 shows the description of the selected variables present in both HMS and MICS 2019, as well as their respective codes in each survey.

Table 2 - Variables present in HMS and MICS 2019

Variable	**HMS Covid-19**	**MICS 2019**
Household head sex	S2 Q5	HL 4
Household head age	S2 Q6	HL 6
Household size	S3 Q5 + Q6 + Q7	HH 48
Number of HH members under 17	S3 Q5 + Q6	HH51 + HH52
Household head education	S3 Q8	ED 5
Sector (urban/rural)	Localidade	HH 6

Source: HMS Covid-19 and MICS 2019

3. **Estimating the model:** the selected characteristics are then used as independent variables to estimate a propensity score model through logistic regression. The dependent variable is participation in the HMS. Thus, the HMS and MICS households are listed according to their propensity score, that is, according to the probability of participating in the HMS given their characteristics. After some tests performed with the variables present in table 2, the final specification of the logistic regression was defined as:

$$logit(p_i) = \beta_0 + \beta_1 male_i + \beta_2 hhsize_i + \beta_3 DepRat_i + \beta_4 hhsize_i^2 + \beta_5 hhheduc_i \\ + \beta_6 urb_i \times distr_i$$

Where p represents the probability of participation in the HMS given the characteristics of the household[1]. $male$ indicates whether the head of the household is male, $hhsize$ is the number of family members, $DepRat$ represents the dependency ratio, $hhheduc$ is a dummy variable for the household head education, and $urb \times distr$ is the interaction between the dummy variables that indicate whether the family lives in urban territory and the district in which they reside. All information is at the household level (i). The results of the model's estimation can be seen Annex 1.

[1] In a more formal way, we have $\Pr(Y_i = y \mid x_{1_i}, \dots , x_{n_i}) = \begin{cases} p_i & if\ y = 1 \\ 1 - p_i & if\ y = 0 \end{cases}$. Where Y_i assumes a value of 1 if the family participates in the HMS, or 0 otherwise. The n independent variables are represented here by x_{ni}.

4. **Ranking of households:** after calculating the propensity score for each household, the families of both surveys are ranked according to that score. The ranking is then divided into subgroups (quintiles) that contain observations from both MICS 2019 and HMS.

5. **Calculating the adjustment factor:** the adjustment factor is calculated using the ratio between the sum of the MICS 2019 sample weights for each quintile and the total MICS 2019 sample weights, on the ratio between the sum of the HMS sample weights for each quintile quintile and the total of the HMS sample weights. More formally, the adjustment factor (f) is calculated using the following equation:

$$f_q = \frac{\sum_{k\in(d_q^m)} weight_k^m / \sum_{k\in(d^m)} weight_k^m}{\sum_{k\in(d_q^h)} weight_k^h / \sum_{k\in(d^h)} weight_k^h}$$

   Where q represents the quintile, d the households interviewed, m the MICS 2019 survey, h the HMS survey, and $weight$is the sample weight. Thus, f_q represents the adjustment factor calculated for each quintile, d_q^m are the households in the MICS 2019 sample for each quintile, d_q^h are the households in the HMS sample for each quintile.

6. **Applying the adjustment factor:** after calculating the adjustment factor, the sample weight of the HMS observations is multiplied by the factor obtaining the adjusted sample weight.

2.2.3 Results

Table 3 shows a comparison between HMS statistics calculated with and without adjusted weights and MICS 2019 statistics.

Table 3 - Comparison between HMS and MICS 2019 statistics

	Percentages		**% of households residing in each district**				
	Male household head	Urban	Água Grande	Mé-Zochi	Noroeste	Sudeste	RAP
MICS 2019	58%	66%	36%	25%	21%	14%	4%
HMS without adjustment	69%	60%	15%	15%	31%	24%	15%
HMS with adjustment	57%	61%	27%	22%	20%	19%	12%

	Averages		**% of heads of household at each level of education**			
	Household size	# of children	Pré-escola/None	Básico	Secundário	Superior
MICS 2019	4.06	2.02	8%	51%	35%	6%

HMS without adjustment	4.60	2.30	1%	54%	38%	7%
HMS with adjustment	3.97	1.93	4%	56%	34%	5%

Source: MICS 2019 and HMS

For all calculated statistics, the values after adjustment are closer to the MICS 2019 survey used as a reference. Despite mitigating the sample's biases, it is still possible to notice differences such as, for example, the greater representativeness of households in the Autonomous Region of Príncipe in the HMS and the underrepresentation of households in which the head of household attended the preschool.

To verify whether the calculated averages are statistically different, an Adjusted Wald test with the null hypothesis that the MICS average is equal to the HMS average was performed. Table 4 shows the results obtained, rejecting the null hypothesis for cases in which the p-value (Prob> F) is greater than 10%.

Table 4 - Adjusted Wald test for the comparison of means between MICS 2019 and HMS

Variable	F (1, 4450)	p-value	Null hypothesis
Male household head	0.46	0.4999	Do not reject
Urban	6.02	0.0142	Reject
Household size	1.11	0.2925	Do not reject
# of children	1.65	0.1994	Do not reject
Água Grande	15.28	0.0001	Reject
Mé-Zochi	2.58	0.1085	Do not reject
Noroeste	0.27	0.6038	Do not reject
Sudeste	13.05	0.0003	Reject
RAP	40.84	0	Reject
Pré-escola/None	8.09	0.0045	Reject
Básico	6.3	0.0121	Reject
Secundário	0.08	0.7734	Do not reject
Superior	2.26	0.1332	Do not reject

The results show that there is a statistically significant difference for the means of the variables of urban area, districts of Água Grande, Southeast and RAP, and heads of household with up to complete basic education.

With the results presented, it is possible to conclude that the process of adjusting the sample weights made the HMS statistics to be closer to the MICS 2019 benchmark at the national level, mitigating potential biases in the sample of the survey conducted by telephone. However, it is important to note that there are still statistically significant differences for some variables, which must be taken into account when using HMS data for statistical analysis.

3. Field Work

3.1 Organization of the Fieldwork

Personnel were selected from the pool of INE interviewers and trained to use CSPro platform and conduct phone surveys. Data were collected by these trained INE interviewers who individually made phone calls from their respective homes. During the preparation and data collection exercise for the first and second round, interviewers were not allowed to be in the office. Therefore, all interviews were conducted from interviewers' homes.

3.2 Gift to Households

As a show of appreciation for the households' participation, all households that gave consent to be interviewed, were transferred 25 dobras credit to their phones (even if their interviews are only partially completed).

3.3 Pre-loaded Information

Basic information on every household was pre-loaded in the CSPro assignments for each interviewer. The information was pre-loaded to (1) assist interviewers in calling and identifying the household and (2) ensure that each pre-loaded person is properly addressed and easily matched to the most recent interviews. Basic household information (location, names, phone number, etc.) was pre-loaded. The list of individuals from the previous interview and their basic characteristics were uploaded. This helped maintain the panel of individuals and ensured the status of each individual in the subsequent round of the survey.

3.4 Respondents

Each round of the STP COVID-19 HMS has ONE RESPONDENT per household. The respondent was the household head or a knowledgeable adult household member. The respondent must be a member of the household. Unlike many other household surveys, interviewers were not expected to seek out other household members to provide their own information. The respondent may still consult with other household members as needed to respond to the questions, including to provide all the necessary information on each household member.

Interviewers were instructed to make every effort to reach the same respondent in subsequent rounds of the survey, in order to maintain the consistency of the information collected. However, in cases where the previous respondent was not available, interviewers would identify another knowledgeable adult household member to interview.

4. Data Management

4.1 Census and Survey Processing System (CSPro)

The STP COVID-19 HMS exercise was conducted using Computer Assisted Telephone Interview (CATI) techniques. The household questionnaire was implemented using the CSPro. Overall, implementation was successful, as it allowed for timely availability of the data from completed interviews.

4.2 Data Cleaning

The data cleaning process was done in two main stages. The first stage was to ensure proper quality control during the fieldwork. This was achieved in part by incorporating validation and consistency checks into the CSPro application used for the data collection and designed to highlight many of the errors that occurred during the fieldwork.

The second stage of cleaning involved a comprehensive review of the final raw data following the first stage cleaning. Every variable was examined individually for (1) consistency with other sections and variables, (2) out of range responses, and (3) formatting. Some minor errors remain in the data where the diagnosis and/or solution were unclear to the data cleaning team.

4.3 Confidential Information

For purposes of maintaining the confidentiality of the data, all names, phone numbers, and addresses have been removed from the datasets.

5. Using the Data

5.1 File Structure

The data should always be used in conjunction with the questionnaire and the interviewer's instruction manual. Where there are no issues of confidentiality, all the variables from the questionnaire have been included in the data sets. The naming of the variables is done using the section number and the question number. For example, the variable s3q8 represents question 8 in section 3 of the questionnaire. In some cases, there is an additional variable which contains the "other specify" information that was written in the questionnaire. This variable will be indicated with an "esp" attached to the variable name such as s3q8_esp containing the "other specify" information for the variable s3q8.

Every effort was made to keep question numbers (and thus variable names) as consistent as possible through different rounds of the survey. If questions were dropped in previous round, the numbering was preserved. If questions were added in the middle of a section, a letter was added to the question number at that space in the sequence (e.g. if added before question 2, the question number would be 2a). This was done to make utilization of the data sets across the rounds as consistent as possible.

5.2 Merging Datasets

Due to the limitations of the survey, it is not possible to merge dataset from different rounds.

6. Survey Rounds

6.1 Round 1

6.1.1 Overview

The STP COVID-19 HMS Round 1 was administered between July 26 and Aug 08, 2020. All 1,400 households selected from the MICS 2019 sample were contacted, with 1,025 of those being fully successfully interviewed. With support from the United Nations, the first round survey was expanded to include a questionnaire aimed at informal businesses.

6.1.2 Weights

The weights can be found in both household-level data files ("hms_stp_hh" and hms_stp_informal"). The variable name is hhweight.

6.1.3 The Survey Instruments

The STP COVID-19 HMS Round 1 consists of two questionnaires. The Household Questionnaire was administered to all households in the sample. The Informal Businesses Questionnaire was administered to households holding informal businesses.

Household Questionnaire: provides information on demographics; knowledge regarding the spread of COVID-19; behavior and social distancing; access to basic services; employment; income loss; and food security.

Informal Businesses Questionnaire: provides information on the impacts of COVID-19 on the household informal business regarding impacts on the labor force; profit and sales; production; how the pandemic impacted the business; and plans for the future.

The contents of both questionnaires are outlined below.

Table 5: STP COVID-19 HMS Round 1 Household Questionnaire

Section	**Topic**	**Description**
Cover	Cover	Household identifiers and enumerator identifiers.
1	Interview information and phone number roster	Roster of call attempts, result and respondent of call attempt, interview consent, date and time of call back, roster of phone numbers, the information of the person that the listed phone number belongs to.
2	Household Basic Information	Roster of members of the household, *and household head gender,* reason for joining the household if new, and reason for leaving the household if left.
3	Household Members	Number of children (0-5 ys old), number of children (6-17), number of adults, and household head education.
4	Knowledge Regarding the Spread of COVID-19	Knowledge of coronavirus, measures to reduce the risk of contracting coronavirus, steps taken by government to curb the spread of coronavirus.

5	Behaviour and Social Distancing	Behavior adopted to prevent infection by COVID-19 (handwashing and social distancing).
6	Access to Basic Services	Household's access to medicine, soap, cleaning supplies, staple food, medical treatment, reason for not being able to access the services, education or learning activities of children at home.
7	Employment	Status and information of income-generating activities (wage work, family business and farming), reason for stopped working, reason for not able to perform activities as usual, and reason for reduced revenue from family business
8	Food Security	Household's food security status during the last 30 days
9	Income Loss	Household's sources of livelihood and their status since the beginning of the outbreak.
10	Interview Results	Result of the interview.

Table 6: STP COVID-19 HMS Round 1 Informal Businesses Questionnaire

Section	Topic	Description
0	Business ID and current situtation	Businesses identification, activity sector, COVID-19 impacts.
1	Labor Force	COVID-19 impacts on labor force.
2	Economic Impact	COVID-19 impacts on sales, revenue, profit, costs, and production.
3	Impact transmission channels	Ways in which the outbreak of COVID-19 impacted the business.
4	Government actions to mitigate COVID-19 impacts	Knowledge and relevance for the business of government social programs aimed to mitigate COVID-19 impacts.
5	Suggestions for new economic stimulus measures	Suggestions for new economic stimulus measures that would be helpful to the business.
6	Adapting the business to a new reality	Measures taken to keep the informal business running.
7	Solidarity mechanisms	Presence of solidarity activities during the pandemic period.
8	Formality	Desire to formalize the business and barriers to formalization.

6.1.4 Description of Datasets

Table 7 shows the datasets files and their description.

Table 7: Dataset files description

Dataset Filename	Description
hms_stp_hh.dta	Round 1 Household
hms_stp_informal.dta	Round 1 Informal Businesses

6.2 Round 2

6.2.1 Overview

The STP COVID-19 HMS Round 2 was administered between January 28 and February 4, 2021. All 1,025 households from the round 1 sample were contacted, with 889 of those being fully successfully interviewed.

6.2.2 Weights

The weights can be found in the household-level data file ("hms_stp_hh_r2"). The variable name is hhweight.

6.2.3 The Survey Instruments

The STP COVID-19 HMS Round 2 consists of one Household Questionnaire administered to all households in the sample.

Household Questionnaire: provides information on demographics; knowledge regarding the spread of COVID-19; behavior and social distancing; access to basic services; employment; income loss; and food security.

The contents of the questionnaire are outlined below.

Table 8: STP COVID-19 HMS Round 2 Household Questionnaire

Section	Topic	Description
Cover	Cover	Household identifiers and enumerator identifiers.
1	Interview information and phone number roster	Roster of call attempts, result and respondent of call attempt, interview consent, date and time of call back, roster of phone numbers, the information of the person that the listed phone number belongs to.
2	Household Basic Information	Roster of members of the household, *and household head gender,* reason for joining the household if new, and reason for leaving the household if left.
3	Household Members	Number of children (0-5 ys old), number of children (6-17), number of adults, and household head education.
4	Knowledge Regarding the Spread of COVID-19	Knowledge of coronavirus, measures to reduce the risk of contracting coronavirus, steps taken by government to curb the spread of coronavirus.
6	Access to Basic Services	Household's access to medicine, soap, cleaning supplies, staple food, medical treatment, reason for not being able to access the services, education or learning activities of children at home.
7	Employment	Status and information of income-generating activities (wage work, family business and farming), reason for stopped working, reason for not able to perform activities as usual, and reason for reduced revenue from family business
8	Food Security	Household's food security status during the last 30 days

9	Income Loss	Household’s sources of livelihood and their status since the beginning of the outbreak.
10	Interview Results	Result of the interview.
11	Vaccine	Coronavirus vaccine-related information: household willingness to take the vaccine, and main concerns related to it.

6.2.4 Description of Datasets

Table 9 shows the datasets files and their description.

Table 9: Dataset files description

Dataset Filename	**Description**
hms_stp_hh_r2.dta	Round 2 Household

References

Lee, S. (2006). “Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys.” Journal of Official Statistics. 22 (2): 329–349.

Rosenbaum, P. R., and D. B. Rubin. (1983). “The Central Role of the Propensity Score in Observational Studies for Casual Effects.” Biometrika 70 (1): 41-55.

Rosenbaum, P.R., and D.B. Rubin. (1984). “Reducing Bias in Observational Studies using Subclassification on the Propensity Score.” Journal of the American Statistical Association. 79: 516-524.

Schonlau M., A. van Soest, A. Kapteyn, and M. Couper (2006). “Selection Bias in Web Surveys and the Use of Propensity Scores.” RAND Labor and Population Working Paper series 229. RAND Pittsburgh, PA.

Terhanian, G., J. Bremer, R. Smith, and R. Thomas. (2000). Correcting Data from Online Survey for the Effects of Nonrandom Selection and Nonrandom Assignment. Research paper: Harris Interactive.

Annex I

Annex 1 - Result of the logistic regression estimation

Variable	Coefficients
male	0.217***
	(0.0818)
hhsize	0.263***
	(0.0756)
DepRat	-0.383
	(0.236)
Hhsize^2	-1.369**
	(0.642)
1.hhheduc (Básico)	1.582***
	(0.279)
2. hhheduc (Secundário)	1.732***
	(0.282)
3. hhheduc (Superior)	1.979***
	(0.313)
Urb (rural) x Distr (Água Grande)	0
	(0)
Urb (rural) x Distr (Mé-Zóchi)	-0.554***
	(0.187)
Urb (rural) x Distr (Noroeste)	-0.427**
	(0.192)
Urb (rural) x Distr (Sudeste)	-0.0212
	(0.185)
Urb (rural) x Distr (RAP)	-0.389**
	(0.196)
Urb (urbano) x Distr (Água Grande)	-0.656***

	(0.176)
Urb (urbano) x Distr (Mé-Zóchi)	-0.706***
	(0.216)
Urb (urbano) x Distr (Noroeste)	0.301*
	(0.177)
Urb (urbano) x Distr (Sudeste)	-0.427**
	(0.186)
Urb (urbano) x Distr (RAP)	0
	(0)
Constant	-3.283***
	(0.338)
Observations	4,396

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1