



# Institute for Health Metrics and Evaluation

## Data Release Information Sheet

### ***Data Summary***

Dataset name: El Salvador Salud Mesoamérica Initiative Baseline Census and Household Survey 2011

Project name: Salud Mesoamérica Initiative Evaluation

Date of release: February 27, 2019

#### Summary:

The Salud Mesoamérica Initiative (SMI) focuses on reducing inequalities in maternal and child health in Mesoamerica. This dataset is the product of an SMI impact evaluation. It includes results of a baseline household census and baseline household survey conducted in eight departments in El Salvador. The census captured basic demographic characteristics of all usual household occupants and was used to produce a sample of households containing eligible women (ages 15-49) and children (ages 0-59 months) for the household survey. In the household survey, any household heads and all eligible women were interviewed. Information was collected on additional demographic characteristics; healthcare use, access, and expenditures; and perceived quality of key interventions for women of reproductive age and children. Height, weight, and anemia measurements were taken for children under age 5.

#### Relevant publications and visualizations:

Mokdad AH, Colson KE, Zúñiga-Brenes P, Ríos-Zertuche D, Palmisano EB, Alfaro-Porras E, et al. Salud Mesoamérica 2015 Initiative: design, implementation, and baseline findings. *Popul Health Metr.* 2015 Feb 7; 13:3. doi: 10.1186/s12963-015-0034-4.

### **Acknowledgments**

#### Contributing organizations:

- Institute for Health Metrics and Evaluation (IHME)
- UNIMER

Funders:

- Bill and Melinda Gates Foundation (BMGF)
- Carlos Slim Health Institute
- Spanish Agency for International Development Cooperation (AECID)  
Inter-American Development Bank (IDB)

***File Information***

**Data Files**

<b>File Name</b>	<b>Description</b>	<b>Data structure</b>
IHME_SMI_SLV_HHS_2011_C ENSUS_Y2019M02D27 [CSV, DTA]	SLV census	Each row represents one household, unique identifier “seg” + “nhogar”.
IHME_SMI_SLV_HHS_2011_ MOD1_Y2019M02D27 [CSV, DTA]	SLV household module 1: 1 survey per household	Each row represents one household, unique identifier “seg” + “nhogar”.
IHME_SMI_SLV_HHS_2011_ MOD2A_Y2019M02D27 [CSV, DTA]	SLV household module 2a: 1 interview with every eligible woman aged 15-49 years in the household	Each row represents one woman, unique identifier “seg” + “nhogar” + “id_woman”.
IHME_SMI_SLV_HHS_2011_ MOD2B_Y2019M02D27 [CSV, DTA]	SLV household module 2b: 1 interview per pregnancy in the last 5 years for each eligible woman aged 15-49 years in the household	Each row represents one live birth, unique identifier “seg” + “nhogar” + “id_woman” + “lb_num”.
IHME_SMI_SLV_HHS_2011_ MOD2C_Y2019M02D27 [CSV, DTA]	SLV household module 2c: 1 interview with the mother/caregiver of each eligible child aged 0-59 months in the household	Each row represents one child unique identifier “seg” + “nhogar” + “id_kid”.

File Name	Description	Data structure
IHME_SMI_SLV_HHS_2011_MOD3_Y2019M02D27 [CSV, DTA]	SLV household module 3: 1 anthropometric measurement module per eligible child aged 0-59 months in the household	Each row represents one child unique identifier “seg” + “nhogar” + “id_kid”.

## Additional File Information

### Inventory

A file inventory (**IHME\_SMI\_SLV\_HHS\_2011\_FILE\_INVENTORY\_Y2019M02D27.XLSX**) contains a list of all files and information on each. This information includes file name, format (CSV, PDF, etc.), type (data, codebook, questionnaire, or documentation), description, and version date.

### Language

The questionnaires and codebooks for all Belize surveys are in English. The questionnaires and codebooks for the El Salvador Baseline Health Facility Survey are in English. The questionnaire for the El Salvador Baseline Census and Household Survey is in both English and Spanish, while the codebooks are in Spanish. Questionnaires and codebooks for all other surveys are in Spanish. Some questionnaires and codebooks also include indigenous languages.

### Codebooks

These contain variable names, questions (variable labels), numeric values and labels for coded values, and question types. Accents have been removed in SMI codebooks, which affects Spanish and indigenous language translations. Some codebooks, however, contain other special characters that do not display properly if a CSV is opened in Excel. Therefore, codebooks are provided in both the machine-actionable CSV format and, for reference, the human-readable XLSX format.

There are nine major types of questions found in the codebooks: calculated, checkAllItem, checkAllSummary, comment, hidden, poplist, preload, radioGroup, and text. These question types are determined by the survey software program and can be grouped into the following categories:

- **Check all that apply:** The check all that apply questions are labeled as either checkAllItem or checkAllSummary
- **Single response option:** The single response options are labeled as either poplist, preload, or radioGroup. These are defined by the type of table that was used in the DatStat survey.
- **Text response:** The text response is labeled as either text or comment
- **Calculated or pre-populated variables:** Any variable that is created by the survey software system, such as time it takes to complete the survey, is labeled as either calculated or hidden. For the purposes of this study, no birth dates or identifiable information is collected, but internal survey calculations were done based on the date of birth to

determine age in years. Other examples of pre-populated variables include geographic information piped into the survey from external census data.

The codebooks contain the most accurate list of variables asked in the survey. The PDF questionnaires produced do not reflect questions that were hidden from participants and interviewers after the initial survey was published for testing and piloting purposes.

## ***Methodological Statement***

### **Data Collection**

Data collection for the SMI-El Salvador baseline measurement was conducted by UNIMER. All surveys were conducted as a pen-and-paper interview (PAPI). Data entry began shortly after the fieldwork commenced and was completed within one week of the end of data collection. Completed household questionnaires were returned periodically from the field to the central headquarters, where they were entered in batch by experienced data entry personnel with training for this task. Data were entered using the computer software package Excel. All data were entered twice.

Portable scales and stadiometers were used for the anthropometric measurements, and hemoglobin levels were assessed in the field using a portable HemoCue machine. Medically trained personnel (i.e., professional nurses) performed all physical measurements.

The SMI-El Salvador Baseline Household Census, which captures basic demographic characteristics of all usual household occupants, was carried out between March 1, 2011 and June 20, 2011. Data collection for the SMI El Salvador Baseline Household Survey began on March 27, 2011 and was completed on July 8, 2011. To assure completeness of the sample, field staff were instructed to return to selected households up to three times (on different days, and at different times during the day) in an attempt to complete the Household Characteristics Questionnaire, the Maternal and Child Health Questionnaire, and the Physical Measurements Module.

Fifteen data collection teams, each consisting of three interviewers (male and female) were deployed to conduct the SMI Household Census. Eleven data collection teams were used to conduct the SMI Household Survey, each consisting of four female interviewers. Supervisors were responsible for reviewing all questionnaires for quality and consistency prior to departing each segment. There were eight supervisors overseeing the SMI Household Census and SMI Household Survey.

The research protocol was approved by the Internal Review Board of the University of Washington. All data collection instruments and procedures were approved by the National Ethics Committee of the Ministry of Health of El Salvador.

### **Sampling/Population**

The sample for the SMI-El Salvador Baseline Household Survey was designed to provide estimates of the coverage of key health interventions and indicators for a geographic area that approximates the lowest wealth quintile of the population of El Salvador.

The primary administrative units in El Salvador are departments, each subdivided into municipalities. El Salvador has 14 departments, and eight were purposefully selected for SMI-El Salvador – Ahuachapán,

Cabañas, Cuscatlán, La Libertad, La Paz, La Unión, Morazán, and San Vicente. From those eight departments, 14 municipalities were identified the Initiative on the basis of their high concentration of residents in the country's lowest wealth quintile. From these 14 municipalities, a two-stage clustered random sample of eligible households was selected.

### **First-stage sample selection: census segments**

The household survey uses a two-stage random sampling design in order to balance survey administration costs with the ability to make estimates representative of the population in the study area. For the SMI-El Salvador household census, the primary sampling unit (PSU) is the *segmento censal* (census segment) from the 2007 El Salvador Population Census. A representative sample of these clusters ("segments") was randomly selected from a sampling frame of all segments in SMI municipalities with probability proportional to size, where size is measured by the number of occupied households. The sample of segments for the SMI-El Salvador Baseline Household Survey was selected twice (hereafter designated the "primary sample" and the "modified sample") because the scope of the project changed after the primary sample had been selected. In addition, an alternate sample was selected in the event that the survey could not be conducted in the selected segments. Below we describe the selection of the primary, modified, and alternate samples.

The primary sample of 136 intervention clusters (segments) were randomly selected from a total of 322 intervention segments in 10 municipalities which, based on data from the 2007 El Salvador Population Census, contained 32,982 occupied households. After the selection of the primary sample was complete, four additional municipalities, comprised of 201 clusters (segments) containing a total of 18,484 occupied households, based on data from the 2007 El Salvador Population Census, were added to the sampling frame.

We calculated the proportion of households in the sample (consisting of ten original municipalities plus the four new municipalities) that were from the four new municipalities. We used this proportion to determine how many segments should be retained from the primary sample (87) and how many segments should be drawn from the four new municipalities (49) for the modified sample. We used a methodology completely analogous to the original sampling procedure to select the 87 retained segments from the original sample of 136 (using probability proportional to size). We then applied the same procedure to the 201 new segments to select 49 new segments with probability proportional to size. This process resulted in a modified sample of 136 segments for this survey.

In addition, a set of alternate segments was selected using identical methodology, to be surveyed in the event that any of the 136 selected segments could not be surveyed and needed to be replaced for any reason (e.g., security concerns or high proportion of absent households). Indeed, security concerns hindered survey implementation in each of the eight randomly selected segments in the department of Cuscatlán. In one instance, a segment in the municipality of Monte San Juan was replaced completely with an alternate segment from the municipality of Ilobasco (in the department of Cabañas). The other seven randomly selected segments in the department of Cuscatlán were prematurely abandoned before adequate samples could be obtained. The samples from each of these segments were supplemented by households from two additional alternate segments (one from Chiltiupán, La Libertad, and the other from El Sauce, La Unión).

## **Second-stage sample selection: households**

Within each randomly selected cluster, a complete household listing exercise was carried out, enabling the systematic selection of households for participation in the survey, based on household composition. All households in which women aged 15-49 years and/or children aged 0-59 months resided were eligible to be selected for the survey.

In order to achieve the desired sample size of 3,800 households, we sought to complete interviews with residents of 28 randomly selected households in each of the 136 randomly selected segments. We surveyed six randomly-selected households with age-eligible women as residents, and 22 randomly-selected households with age-eligible children as residents. To do so, listings of all households with age-eligible women or children were assembled in random order for each segment. Naturally, there was a substantial degree of overlap between houses listed on the “woman-resident” list and houses listed on the “child-resident” list. In each segment, household selection occurred in sequential order from both lists, until a total 28 households had completed interviews (6 from the list of “woman-resident” households, and 22 from the list of “child-resident” households). In some cases, selected households were absent or declined to participate in the SMI Household Survey. These households were replaced by other households from the same segment, when alternate households with age-eligible residents existed. Because multiple interviewers worked the sample simultaneously, in a handful of instances more than 28 surveys were completed. This occurred in 11 segments, where between 29 and 31 households completed surveys.

All women aged 15-49 years who are members of the selected household are eligible to be interviewed, and all children aged 0-59 months who are members of the selected household are eligible for the physical measurement module. Any household head or other individual knowledgeable about household characteristics and expenditures is permitted to respond to the household characteristics module, while any primary caregiver of a child 0-59 months is eligible to inform for the child health interview module, regardless of sex or age.

## **Weighting**

Survey weights reflect the three-stage cluster sampling design of the study. The primary sampling unit is referred to as the “segment.” The segment is censused, and 28 households with eligible participants selected at random. Within selected households, all women 15-49 years of age and all children 0-59 months of age are selected for participation in the survey. Design weights for households, women and children were calculated according to the inverse probability of selection of the unit. No post-stratification adjustments were made to the weights. Although cluster sampling can improve efficiency when the target population is spread out over a large area, the resultant sample consists of observations that are not completely independent of one another. Estimation using SMI survey data should apply sampling weights in order to represent the population of the study area, and should account for intra-class correlation by specifying clusters and strata when calculating variance.

## **Imputed Variables and/or Constructed Variables – What was Imputed/Constructed and How**

There are no imputed variables in the data. The constructed variables are labeled in the dictionary as “IHME Generated” and consist only of the weights constructed by the research team.

## **Known Data Quality Issues**

### **Sampling errors & Design Effects for Key Indicators**

- Age is registered at the time of the household census, so age in months may represent a lag of up to a month at the day of interview/physical measurements.
- Continuous data (e.g., household expenditures) have not been cleaned or trimmed for outliers.
- There are data entry errors, specifically with the child roster IDs in Modules 2A and 2B, that were entered in the field. IHME had a thorough data verification process and communication system with the field team, although not all originally reported data may align throughout the census to the household.
- Ages in livebirth roster have not been reconciled or corrected except in some cases where the child is under age 5.
- Module 3 had a separate consent process for child anthropometric measurements and was sometimes refused even for children who have an interview module 2C.
- During the cleaning and measurement process for country indicators, some don't know or decline to respond answer values were irreversibly replaced with missing values.
- The El Salvador baseline household survey was captured on paper and sometimes non-standard codes were used for responses of "don't know" or "decline". The values may appear in the data.
- Some households may be missing one or more modules due to refusal or tracking errors in the field.

## **Public Use Dataset Notes**

This is a public use dataset. The data have been de-identified. Variables determined to contain identifiable private information, or potentially identifiable private information, for health facilities, health workers, and/or other individuals have been removed in accordance with IHME's microdata release protocol. The protocol's determination for variables that constitute identifiable private information is based primarily on [HIPAA'S De-identification Standard](#).

No personally identifiable information was collected for this study; however, these data were stripped of comments and information on who conducted the interview. Some variables in the dataset do not contain data, such as date of birth, because this information was not stored on the survey or sent to IHME after data entry.

## ***Additional Information***

### **Terms and Conditions**

<http://www.healthdata.org/about/terms-and-conditions>

### **Contact Information**

To request further information about the Salud Mesoamérica Initiative (SMI), please contact IHME:

Institute for Health Metrics and Evaluation

2301 Fifth Ave., Suite 600

Seattle, WA 98121

USA

Telephone: +1-206-897-2800

Fax: +1-206-897-2899

Email: [data@healthdata.org](mailto:data@healthdata.org)

[www.healthdata.org](http://www.healthdata.org)

These files may be updated periodically, so we appreciate hearing feedback or additional information about how these data are being used.