



Institute for Health Metrics and Evaluation

Data Release Information Sheet

Data Summary

Dataset name: Mexico - Yucatán Under-5 Baseline Household Survey 2020

Date of release: March 3, 2021

Summary:

These data are the product of a collaboration between the Institute for Health Metrics and Evaluation (IHME) and the Universidad Autónoma de Yucatán (UADY). The objective of the project was to improve maternal and child health and the quality of health information in the state of Yucatán, Mexico through assessing the knowledge of alarm signs, and access and utilization of health services, among caregivers of children under 5 years of age. This dataset includes the results of a household census and caregiver interviewer. The population under study includes caregivers of children under 5 in 8 municipalities in Yucatán. In total, data were collected from 2,996 households.

Acknowledgements

Contributing organizations:

- Institute for Health Metrics and Evaluation (IHME)
- Universidad Autónoma de Yucatán (UADY)

Funding institution:

- W.K. Kellogg Foundation

File Inventory

File Name	Description	Version Date
IHME_MEX_YUCATAN_BASELINE_HH_SU RVEY_2020_DATA_Y2021M03D03.CSV	Public use survey dataset- CSV format	March 3, 2021
IHME_MEX_YUCATAN_BASELINE_HH_SU RVEY_2020_CODEBOOK_Y2021M03D03. CSV	Codebook, with descriptions of all variables included in the dataset	March 3, 2021
IHME_MEX_YUCATAN_BASELINE_HH_SU RVEY_2020_QUESTIONNAIRE_Y2021M03 D03.PDF	Questionnaire	March 3, 2021

File Name	Description	Version Date
IHME_MEX_YUCATAN_BASELINE_HH_SU RVEY_2020_INFO_SHEET_Y2021M03D03. PDF	Data release information sheet	March 3, 2021

Additional File Information

Data Structure

The data are structured such that each row represents a household. In households where caregiver interviews were conducted, both census and caregiver interview responses are contained in a single line of data. Portions of the caregiver interview are repeated based on the number of caretakers and children under the age of 5 present within the household.

Language

The questionnaire and codebook are in Spanish.

Codebooks

These contain variable names, question text, numeric values and labels for coded values, and question types. Accents have been removed, which affects Spanish language translations.

There are four major types of questions found in the codebooks:

- Check all that apply: The check all that apply questions are labeled as “select_multiple”
- Single response option: The single response options are labeled as “radioGroup”
- Text response: The text response is labeled as “text”
- 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 “calculated”.

Methodological Statement

Data Collection

This project received approval from the internal review boards from the University of Washington and the Universidad Autónoma de Yucatán as required by the Secretary of Health, to provide authorization to conduct the study.

Data collection was conducted by a team of interviewers from UADY. All surveys were conducted using a computer-assisted personal interview (CAPI). The CAPI was programmed using SurveyCTO and installed onto computer tablets. CAPI supports skip patterns, inter-question answer consistency, and data entry ranges. The aim of introducing CAPI to the field was to reduce survey time by prompting only relevant questions, maintain a logical answering pattern across different questions, decrease data entry errors, and permit rapid data verification.

The interviews were carried out between January and April 2020. Informed consent was obtained prior to all interviews. To ensure that the collected data remains anonymous and confidential, all data from the

interviews and household rosters were systemized in a database using a study identifier that de-identified the data in a way that does not allow users to trace back the identity of study participants.

All data collected by UADY in the field were uploaded onto a server using a secure link. Quality checks were conducted by the IHME team in real time to clarify doubts as data was collected.

Sampling/Population

Eligibility for participation in the caregiver interview included all caregivers of children under 5 years of age in 8 municipalities of Yucatán, Mexico (Bokoba, Buctzotz, Calotmul, Cantamayec, Chapab, Dzan, Sotuta, and Tekanto). Municipalities were selected to be included in the study based on high rates of child mortality, having a Secretary of Health clinic, and lastly, being safe and reasonable for the interview team to reach. Communities within municipalities were selected based on the census track (AGEB) by the UADY team and then a full census was conducted. In households without a child under 5, only a family roster was collected. In households with a child under 5, interviewers identified that child's primary caregiver for an interview immediately following the household roster. The inclusion criteria for caregiver interviews were that the caregiver was 18 years of age or older and had physical and mental capabilities to provide information about the child.

Weighting

The data are unweighted.

Imputed variables and constructed variables

No variables are imputed. The SurveyCTO software automatically generates certain variables for each interview conducted. Additionally, IHME programmed SurveyCTO to generate additional variables, including reconfigurations of responses previously provided in the survey, to facilitate the flow of the survey. Both variable types described above are characterized as having a *question_type* equal to "calculated" in the codebook.

Known Data Quality Issues

Data collection teams experienced elevated levels of refusals to participate in the study during the time of data collection due to multiple high-profile killings of women and girls that occurred in Mexico City. This resulted in increased security measures within many communities across Mexico. Data collection teams collaborated with local authorities to spread awareness of the presence of data collection teams within the community, and response rates improved.

Continuous data have not been cleaned or trimmed for outliers.

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. The date of birth was entered into the survey and an internal calculation was done to provide age.

Additional Information

Terms and Conditions

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

Contact information

To request further information about this dataset, please contact IHME:

Address:

Institute for Health Metrics and Evaluation
Population Health Building/Hans Rosling Center
3980 15th Ave. NE, Seattle, WA 98195 USA
UW Campus Box #351615

Telephone: +1-206-897-2800

Fax: +1-206-897-2899

Email: data@healthdata.org

Website: www.healthdata.org

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