



Institute for Health Metrics and Evaluation

Data Release Information Sheet

Data Summary

Dataset name: Mexico - Yucatán Under-5 Endline 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. The population under study includes caregivers of children under 5 in 8 municipalities in Yucatán. This survey covered topics related to the identification of symptoms for common causes of death, health-care seeking behaviors, and a short series of questions related to COVID-19. In total, responses were collected from 500 respondents.

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_ENDLINE_HH_SURVEY_2020_DATA_Y2021M03D03.CSV	Public use survey dataset- CSV format	March 3, 2021

File Name	Description	Version Date
IHME_MEX_YUCATAN_ENDLINE_HH_SURVEY_2020_CODEBOOK_Y2021M03D03.CSV	Codebook, with descriptions of all variables included in the dataset	March 3, 2021
IHME_MEX_YUCATAN_ENDLINE_HH_SURVEY_2020_QUESTIONNAIRE_Y2021M03D03.PDF	Digital Questionnaire	March 3, 2021
IHME_MEX_YUCATAN_ENDLINE_HH_SURVEY_2020_QUESTIONNAIRE_PAPER_Y2021M03D03.PDF	Paper Questionnaire	March 3, 2021
IHME_MEX_YUCATAN_ENDLINE_HH_SURVEY_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 an individual respondent.

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 working in collaboration with health workers from local health clinics. All surveys were conducted on paper and subsequently entered into an electronic database using the SurveyCTO platform. The SurveyCTO electronic data capture system was programmed by IHME and completed by UADY team members. The data capture system was developed with minimal restrictions on entered data to allow data transcribers the flexibility to enter data exactly as it appeared on the original paper submission. The aim of utilizing a paper-based data collection system was to minimize the need for close contact among survey participants and research staff due to the ongoing threat of COVID-19.

Interviews were conducted between September and November 2020 and entered into the electronic databased in November and December 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. The survey covered topics related to the identification of symptoms for common causes of death, health-care seeking behaviors, and a short series of questions related to COVID-19.

Quality checks were conducted by IHME concurrent to electronic data capture to identify and resolve data quality concerns.

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 during the baseline phase 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. During baseline data collection from January – March 2020, communities within municipalities were selected based on the census track (AGEB) by the UADY team and a full census was conducted. At baseline, a family roster was collected at households without a child under 5; while at households with a child under 5, interviewers identified that child’s primary caregiver for an interview immediately following the household roster.

Caretakers who completed the baseline survey in January – March 2020 were contacted to complete the endline survey. Any additional individuals within the selected communities who were not interviewed at baseline but who acted as a primary caretaker of a child under 5 were also eligible for the endline survey, however these individuals were not actively recruited to participate in the survey.

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

The survey was originally conceived of and designed to be completed electronically on tablets with team members from UADY serving as data enumerators and to be conducted at the household of each participant. Data collection at baseline was conducted as originally planned. As a result of COVID-19 and various government mandates aimed at mitigating the spread of the disease, this methodology and mode for data collection was deemed unsafe for endline data collection. Endline data was instead collected using an abbreviated, paper-based survey at local health clinics with health facility personnel to reduce the risk of contracting and spreading COVID-19. Changes to data collection mode and methods between baseline and endline surveys may affect survey responses.

Endline surveys were originally designed to be administered approximately 6 months after the implementation of interventions in order to allow for health-care seeking events to occur in the period between intervention and evaluation. Due to delays in the implementation of interventions resulting from COVID-19, endline data collection was completed almost immediately after the implementation of interventions. This adjusted timeframe limits the ability to attribute changes in healthcare seeking behavior to intervention activities.

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.