Mexican Health and Aging Study MHAS 2012

Sample design

Estudio Nacional de
Salud y Envejecimiento
en México

INEGI

INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

Sample design of the Mexican Health and Aging Study (MHAS) 2012

Presentation

The Instituto Nacional de Estadística y Geografía (INEGI), in the framework of the powers conferred by the Ley del Sistema Nacional de Información Estadística y Geográfica, develops and implements strategies aimed to consolidate the Sistemas Nacionales de Información Estadística y Geográfica (SNIEG), among which is the documentation of the different projects carried out.

In this context the Sample Design of the Mexican Health and Aging Study (MHAS) 2012 is presented with the purpose of providing an overview of the methodology used in the design and development of the interview.

Thus, INEGI offers users a document that shows, in summary, the different aspects of statistical design of the Project.

Also, with these types of documents it is possible to contribute to the clarity of the process of collecting statistical data to provide an example that will allow improving the design and implementation of future projects in this field.

Index

Introduction

The Mexican Health and Aging Study (MHAS) 2012, is a longitudinal project that updates the statistical information collected in the 2001 and 2003 surveys to the population 50 and older in Mexico, with the purpose of evaluating the aging process, the impact of disease and disability in performing activities regardless of the nature of these.

The survey provides demographic data of people, employment, housing, use of time, as well as on their health, disability, migration, heritage and socioeconomic status, among other subjects.

In order to frame the main methodological characteristics of the study, the MHAS 2012 sample design was created, which addresses conceptual topics, general characteristics and statistical design used to develop the survey.

The sample design is organized into three parts; the first addresses the background and legal framework; the second addresses objectives, conceptual topics and methodological bases; and the third part is devoted to the statistical design. Tables with indicators of stratification by study area and the distribution of the sample by entity are presented in the Annex.

Therefore, INEGI disclosed in summary form, the methodology and contributes to the transparency of the process of generating statistical information.

1. Background[1]

The aging process and the conditions of life and health of the adult population are topics of significant relevance in recent years, because they constitute challenges the country must face and address urgently. In this light, it is essential to have information to permit sizing the phenomenon, to observe it from different angles, and appreciate their consequences.

In this regard, there have been various studies on the topic, and despite their limitations or lack of disseminating results; they provide a platform to reach better understanding.

The Consejo Nacional de Población conducted the Encuesta Nacional Sociodemográfica sobre el Envejecimiento in 1994, which provides estimates of functionality in older adults. The project covered a sample of 5,000 individuals and its results have national representation.

The Universidad de las Naciones Unidas in 1995 undertook a multinational comparative study on nutrition in older people called Cross-cultural Research on Nutrition of Older Subjects, which corresponds to a cross-sectional survey on the nutritional status of older people, compared to middle aged individuals in three communities: urban, suburban and rural. It provides information on nutritional state and its relationship with general health, functional status and socioeconomic conditions.

In order to describe the health status of this population, assess both levels of functional limitation, and the use and accessibility to medical services, the Pan American Health Organization conducted the Encuesta sobre la Salud y el Bienestar en el Envejecimiento (SABE). The study was conducted in 1999 and included a total of 8000 individuals in seven urban areas in seven Latin American countries: Argentina, Barbados, Brazil, Chile, Cuba, Mexico and Uruguay.

It has also conducted research on the subject in more specific areas. Such is the case of the study Envejecimiento Poblacional en el Instituto Mexicano del Seguro Social, held in Mexico City in 1995, with a sample of 5,500 beneficiaries.

The evaluation on quality of life in institutions for the elderly in the Mexico City was held in 1995 under the auspices of the Pan American Health Organization, in order to observe the quality of care provided in long-term care institutions for the elderly. The study was then repeated using the experience gained and the methodology proposed by the French National Foundation for Gerontology.

The Encuesta sobre Envejecimiento Cognoscitivo en la Ciudad de México was administered in 1999 to a sample of 4,047 individuals. The project was unique in emphasizing dementia, cognitive impairment and depression, their implication on functionality, and their relationship with health status and sociodemographic factors.

---

[1] Gutiérrez Robledo, Luis Miguel. La salud del anciano en México y la nueva epidemiología del envejecimiento.

The Mexican Health and Aging Study (MHAS) was administered for the first time in 2001, through the collaboration of the Instituto Nacional de Estadística, Geografía e Informática (INEGI) and researchers from the universities of Pennsylvania, Maryland and Wisconsin in United States of America (U.S.), with the goal of collecting information about the process of aging of the population 50 years and over, as well as on their socioeconomic conditions, health and disability.

Later in 2003, the second wave of the MHAS was carried out to monitor the information from the previous survey. The study design allowed us to identify the predominant features and characteristics of the sample and how they evolved.

1.1 Legal framework

The University of Texas Medical Branch is a non-profit institution dedicated to provide higher education, and has legal authorization of the National Institutes of Health under the Department of Health and Human Services of the U.S. to collaborate with INEGI in conducting MHAS 2012.

INEGI, in compliance to the mandate of the Ley del Sistema Nacional de Información Estadística y Geográfica (LSNIEG), which gives them the character of normative institution and coordinator of the SNIEG, participated in the design of the survey and was in charge in collecting and processing the survey information.

2. Methodological and conceptual basis

In this section we present the objectives and a summary of the design aspects such as like thematic and conceptual framework, data collection instruments, and methodological references supporting them.

2.1 Objectives
General
Update statistical information gathered in previous surveys (2001-2003) of the population 50 and older in Mexico, with urban and rural representation, to evaluate the aging process, and the impact of disease and disability in performing daily activities.

Specific
- Update sociodemographic characteristics of the selected persons, their spouses and other residents in the households of respondents in 2003, and collect data from new households selected in the sample.

- Collect supplementary demographic data for selected individuals, including migration history.

- Update or collect sociodemographic characteristics of the non-resident children and deceased children.

- Collect data to determine the overall health of the interviewed population and identify health services utilization as well as the costs incurred in obtaining such services. Data to determine the overall health of the population interviewed and identify health services to those uses, as well as the costs incurred in obtaining such services.

- Apply a series of cognitive exercises that enable researchers to determine the level of memory and mental health of people.

- Learn about the survival of the parents of the interviewees, the relationship with them and the help they are given.

- Gather information needed to quantify the amount and identify the type of help that is given and received from their children.

- Gather information to assess the state of functionality and assistance the interviewed person receives to perform their activities of daily living. Generate information to assess the state of functionality and assistance it receives the respondent to develop their daily activities.

- Provide information on work history, current employment status, and characteristics of their current job.

- Obtain characteristics, property status, and the estimated value of their properties.

- Gather information on income received from employment, pensions or other sources, and the value of goods estate and capital.

2.2 Conceptual aspects

The conceptual design of the MHAS 2012 was the result of several exchanges between the University of Texas Medical Branch and INEGI, as well as the implementation of the pilot test, with the primary purpose of strengthening procedures to capture quality information to meet project objectives and cover the wide range of topics related to aging, health, disability, family networks and financial support, among others.

In general, the original conceptual framework was maintained in order to ensure comparability with the results of previous projects and thus the possibility of analyzing the evolution of the topics object of the study.

Study topics

Presented below are topics addressed in this survey. The level of detail differs by questionnaire.

- Housing characteristics. Type of housing, building materials and other indicators of housing quality, property conditions and availability of consumer durables.

- Sociodemographic characteristics of the selected individual. For example, date and place of birth, education, marital status, fertility, unions or previous marriages, migration history. These variables are used to understand the context of life of the interviewee.

- Economic data. Earnings and other sources of income, including pensions and government subsidies, type and value of both monetary and real estate.

- Migration experiences. From the selected individual, his/her parents, siblings and children.

- Family. List of all children, including those who do not reside with the older adult, and deceased children; and for each one, their demographic data, indicators that summarize current and childhood health, education, and current economic situation.

- Self-assessments. Self-report of general health, personal review of individual economic status, level of control over decisions and level of family support.

- Health measures. Reports of conditions, symptoms, functional status, lifestyle (e.g., history of smoking, alcohol consumption and exercise), depression, illness, reading performance and cognitive state.

- Employment. Issues such as employment history, occupation, position at work and current workplace.

- Transfers. Help in cash or kind and time provided to their parents or received from their children.

- Widowhood. For individuals in this status, questions about changes in their economic status, employment and living arrangements as a result of the death of a spouse.

Conceptual topics

The following are the basic concepts essential to understand both the conceptual design and the collecting procedures:

Permanently absent. A person who ceased to be a resident of the house by being absent for more than three months, regardless of the reason for his absence.

Temporarily absent. A person that at the time of visit is not in the house, but intends to return during the collection period or before three months, regardless of the reason for his absence.

Unique household ID (CUNICAH). Is the code that identifies households in the sample and allows for the grouping of new and original individuals by households (2001 sample).

Future contact. A person who is directly related with the study subject, does not live in the same dwelling, but can provide information in case of change of address within two years. It can be: mother, father, brother(s), sister(s), aunt(s) or uncle(s).

Core questionnaire (CQ). Collection instrument used to obtain information from selected study subjects.

Next-of-Kin questionnaire (NQ). Collection instrument used to obtain information about the deceased subject, through a proxy interview.

Proxy questionnaire. Collection instrument used to obtain information from study subjects, through a proxy informant who knows details of the selected subject.

Geographical residence. According to the technical definition, is the space within a locality or refers to a communication path that occupies a property (building or land) where one or more persons or economic units can be established in order to comply with its obligations or rights.

Follow-up interview. It is the conversation or dialogue between the interviewer and MHAS informants that were visited previously during data collection in 2001 and 2003.

Proxy interview. Corresponds to the conversation or dialogue established between the MHAS interviewer and the informant who knows the details of the subject of study, which it is not able to answer the questionnaire because of some health impairment due to an accident, illness, advanced age, or because of language or temporary absence.

Household. Consists in a group of one or more persons, with or without kinship ties, which usually reside in the same private dwelling and live on each other's income for sustenance.

Proxy informant. Person 18 years or older who is familiar with information of the study subject, which is not able to answer the questionnaire due to accident or illness, advanced age, language or temporary absence.

Informant of deceased subject. Person who had frequent contact with the subject of study before dying (family member or guardian) who can provide information on personal situations and cause of death.

Person Number (NP). Is the code that identifies subjects in the sample. The code consists of three digits and can differentiate persons originally selected (sample 2001), spouses of individuals originally selected, and subsequent partners of both sexes. The originally selected subject (baseline) is always assigned the number 010 as their identification code.

Relationship. Link with household members to the head of household, whether by blood, marriage, family relation or adoption.

Partner or spouse. Person recognized as a life partner, regardless of whether they live in the same house and their union is legalized or not.

Pre-loaded. Information captured by the MHAS 2001 and 2003 served as a reference during fieldwork in 2012.

First respondent (contact). Person by whom the interviewer can get permission to visit the dwelling and can inform on the situation on subjects or their new home.

Report on pre-loaded information. Function available in the utilities menu of the MHAS Collection System 2012. Includes the dataset collected in 2001 and 2003 to build a follow-up interview when using paper questionnaires.

Permanent resident. Person who normally lives in the dwelling, where they usually sleep, prepare food, eat and are protected from the environment, and therefore recognized it as their residence.

Sub-household. It is the code that identifies the new homes that result from separations or new unions of study subjects from the 2001 and 2003 sample and are followed in the survey.

Deceased subject of study. Person who is part of the 2001-2003 sample of the study, selected individual or spouse, and that when applying introductory questions in 2012 there was the indication that the subject died.

Adequate informant. A person 18 years or older, resident of the household, who knows the sociodemographic information of the residents.

2.3 Survey design

The project was carried out at the request of the University of Texas Medical Branch (UTMB), institution responsible of the conceptual design and the definition of the questionnaires; tasks undertaken under the criterion of ensuring the conceptual comparisons with MHAS 2001 and 2003 results, since this is a longitudinal study that aims to establish the monitoring of study subjects to provide information to analyze the evolution of health conditions and aging.

From the experience learned in previous surveys, as well as the considerations that arise from the current environment, the results of the pilot test and other information required from UTMB, some adjustments were made that contributed to enrich the collection instruments.

While INEGI made comments to the questionnaires, the institution fundamental tasks were the design and development of the processes of collecting and processing the data.

Five instruments were used for different purposes, to capture specific characteristics.

Introductory questions questionnaire

The questionnaire consists of two sections, each of them is a separate instrument used to collect specific data.

The first section of the questionnaire was design to identify if subjects were still living and to verify current residence and marital status in eligible households with one follow-up subject or one selected subject from the new sample. The second section was applied to households with couples from the follow-up sample, and allowed to identify survival status, residence and possible changes in marital status.

Introductory questions questionnaire sections

Sections		Number of questions
	Total	**33**
I	Household with one follow-up subject or subjects from the new sample	9
II	Household with couples (follow-person and partner or spouse)	24

The first section consists of six sub-sections, while the second has 16. The introductory questions questionnaire is the instrument used to collect data necessary to establish a broad view of the study subjects and to identify the specific subsequent instruments used during the interview.

Core questionnaire

The questionnaire contains two household roster cards, one for follow-up subjects and the other new individuals; a format for first interview (control format), a format to record the results of the interview, 13 sections and a format to record information for future contact.

Core questionnaire sections

Sections		Number of questions
	Total	**57**
TRH	Household roster card for follow-up interview	26
B	Non-resident children for follow-up interview	31
	Total	**59**
TRH	Household roster card for new person interview	23
B	Non-resident children for new person interview	36
	Total	**517**
A	Demographic data for follow-up person	40
AA	Demographic data for new person	43
C	Health	75
D	Control and health services	35
F	Parents and help to parents	59
G	Help and children	35

H	Functionality and help	36
I	Employment	31
J	Housing	37
K	Pension, income and assets	118
SA	Surviving widow(er)	8

Major events and cognitive exercises questionnaire

This questionnaire is composed of two different instruments given at different times

Section topics of questionnaire of major events and cognitive exercises

Sections		Number of questions
	Total	23
I	Major events	10
II	Cognitive exercises	13

Next-of-kin questionnaire

The questionnaire consists of an interview format for the deceased (control format), 9 sections to capture information and one for general observations about the interview.

Next-of-Kin questionnaire sections

Sections		Number of questions
	Total	**118**
SA	Major events	18
SB	Temporary living arrangements	8
SC	Health	31
SD	Health services	14
SE	Cognitive condition	10
SG	Help and children	11
SH	Functionality and help	8
SI	Employment	6
SJ	Cognitive exercises	12

Questionnaire for proxy interviews

The questionnaire contains two household roster cards to record residents, one for follow-up subjects and the other for new individuals, an interview format by proxy respondent (control format), a format to

record the results of the interview, 11 sections and a format to record information for future contact.

Introductory questions questionnaire sections

Sections		Number of questions
	Total	**57**
TRH	Household roster card for follow-up interview	26
B	Nonresident children for follow-up interview	31
	Total	**59**
TRH	Household roster card resident for new person interview	23
B	Nonresident children of new person interview	36
	Total	**371**
A	Demographic data for follow-up person	16
AA	Demographic data for new person	21
C	Health	47
D	Control and health services	15
PC	Proxy cognitive condition	63
F	Parents and help to parents	22
G	Children and help to children	17
H	Functionality and help	11
I	Employment	22
J	Housing	33
K	Pension, income and assets	104

2.4 Methodological framework

In order to highlight the main topics of the survey, we present the following methodological framework:

Sampling Unit. This includes people 50 or older and their spouse or partner(s), if any, regardless of age.

Sampling frame. The follow-up sample corresponding to those selected in the 2001 MHAS wave and their spouses, regardless of their age, as well as new potential partners that have been added to the study in 2003. The new sample corresponds to persons 50 to 60 years and their partners, selected from the National Occupation and Employment Survey (ENOE) in the second quarter of 2012.

Geographic areas. They survey is designed to provide nationally representative results.
Reference period. The survey asks about events and situations occurring in different periods, some highlighted are:
- Date of interview is considered in almost all sections of the questionnaire

- For the last two years, themes on health, parents and help to parents, children and help to children
- For the past 10 years for the health, children and help to children, employment and major events

Data collection. Data collection was conducted from October 1 to November 23, 2012.

Sample size. The total sample of MHAS 2012 is totaled at 20,542 and is broken down into two types:
- The first was interviewed in 2001 or 2003 and followed with a total of 14, 283 individuals
- The second is an additional sample obtained from ENOE of the second quarter of 2012 sample with a total of 6,259 persons.

In order to study the anthropometric characteristics of the study population, a subsample of 2,475 persons was selected.

Data collection method. The data was obtained by direct interview using a printed questionnaire and an electronic version, using a mini laptop and given to the respondent or a proxy, if the study subject had any impediment preventing them from completing the questionnaire or if the subject died. The data collection instruments are structured with questions based on respondent answers, where some answers are unavailable for some questions and present in others.

Study subjects. The study focuses on:
- People 50 or more years of age selected in 2001 and their spouses or partners identified in the 2001 survey or those followed up in 2003, the deceased, those still alive and residing in a private or community dwelling in Mexico
- People 50 to 60 years of age, selected in 2012 and their spouses or couples living in the same dwelling in Mexico

In cases where the selected person had a encumbrance, such as illness, disability or has died, the proxy informant who had frequent contact and was close to the deceased study subject provided information for the survey.

3. Statistical design

In selecting the sample the following were considered: a sufficient size was used to make estimates for the target population, a sampling frame and the construction and evaluation of estimates to broaden the use of information from the field data.

3.1 Sampling frame

The sampling frame was carried out as follows:

- Follow-up sample. Includes study subjects interviewed in the survey in 2001 and 2003
- Additional sample. This lists the houses where at least one person between 50 and 60 years of age lived in and where identified from the National Survey of Occupation and Employment of the second quarter of 2012.

3.2 Sample size

The total sample size for the study is 20,542 individuals. Of these, 14,283 are follow-up subjects from the 2003 survey. Further, 6,259 individuals come from the new sample between the ages of 50 to 60 years of age at the second quarter of 2012 The sample distribution is presented in Table 1 by state.

Subsample for biomarkers

Biometric testing was implemented in all subjects from ██████████████████████████████████
The distribution is presented in Table 2.

3.3 Sample selection

The sample selection involved three stages. In the first stage, the primary sampling units (PSU) are selected. Second, within each primary sampling unit, households are selected. In the third stage, persons in the selected households are selected for the study.

For the follow-up sample, the probability of selecting a person of interest as well as its weights, are given by the calculations of MHAS 2001. For the new sample selection, the procedure is described below:

Households' selection

From the N_{eh} PSU identified in the ENOE 2012, n_{eh} PSU were randomly selected within the population between 50 and 60 years of age.

Therefore, the probability of selecting the k-th household in the i-th PSU in the h-th stratum in the e-th state is:

$$P_{ehik} = P_{ehik}^{ENOE} \frac{n_{eh}}{N_{eh}}$$

Its weight is:

$$F_{ehik} = F_{ehik}^{ENOE} \frac{N_{eh}}{n_{eh}}$$

Where:

P_{ehik}^{EN0E} = is the probability of selecting the k-th housing, in the i-th PSU of the h-th stratum and the e-th state, from the ENOE 2012.

n_{eh} = is the number of households with a person between 50 and 60 years of age selected in the h-th stratum, in the e-th state in MHAS 2012.

N_{eh} = is the number of households with a person between 50 and 60 years old, in the h-th stratum, in the e-th state, identified by the ENOE 2012.

F_{ehik}^{ENOE} = is the expansion factor of the k-th housing, the i-th PSU of the h-th stratum, and the e-th entity according to the ENOE 2012.

Subject selection

Once a household is selected, the method of selection of the individual is the following:
In each selected household selected, one person between 50 to 60 years of age is randomly selected. If the person selected has a partner, that person is also selected as a study subject.

To determine the probability of selection, three cases are explained below:

Case 1. Individuals between 50 to 60 years old, whose partner is 50 to 60 years, both living in the household.

In the selected household, the probability that the ℓ-th person between 50 and 60 years old with a partner in the same conditions is selected, is that one is directly selected or indirectly selected through their partner, i.e.:

$$P_{ehik\ell}=P_{ehik}\left[\frac{1}{N_{ehik}}+\frac{1}{N_{ehik}}\right]=\frac{2}{N_{ehik}}P_{ehik}$$

Its weight is given by:

$$F_{ehik\ell}=\frac{N_{ehik}}{2}F_{ehik}$$

Where:

P_{ehik} =   is the probability of selecting the k-th housing, in the i-th PSU of the h-th stratum and the e-th state, from the ENOE 2012.

N_{ehik} =   is the number of people between 50 and 60 years old, in the k-th property in the i-th PSU, in the h-th stratum and the e-th state..

F_{ehik} =   is the expansion factor for the k-th housing, the i-th PSU of the h-th stratum, and the e-th entity.

Case 2. Individuals between 50 and 60 years of age without a partner in the same household.

In this case the probability of selection is given by:

$$P_{ehik\ell}=P_{ehik}\left(\frac{1}{N_{ehik}}\right)$$

Its weight is:

$$F_{ehik\ell}=N_{ehik}F_{ehik}$$

Where:

P_{ehik} =   is the probability of selecting the k-th housing, in the i-th PSU of the h-th stratum and the e-th state, from the ENOE 2012.

N_{ehik} =   is the number of people between 50 and 60 years old, in the k-th property in the i-th PSU, in the h-th stratum and the e-th state..

F_{ehik} =   is the expansion factor for the k-th housing, the i-th PSU of the h-th stratum, and the e-th entity.

Case 3. Individuals who are not between 50 and 60 years old, partners of a person between 50 and 60 years, and living in the same household.

$$P_{ehik\ell}=P_{ehik}$$

Its weight is:

$$F_{ehik\ell}=F_{ehik}$$

Where:

P_{ehik} = is the probability of selecting the k-th housing, in the i-th PSU of the h-th stratum and the e-th state, from the ENOE 2012.

F_{ehik} = is the expansion factor for the k-th housing, the i-th PSU of the h-th stratum, and the e-th entity.

3.4 Adjustments to the weights

The weights constructed according to the process described above are set for the following adjustments:

Adjustments for non-response

The weights calculated for the follow-up sample and the additional sample are adjusted for non-response, as follows.

A) New sample

The non-response adjustment attributed to the informant was calculated at the PSU level using the following expression:

$$F'_{ehik\ell}=F_{ehik\ell}\,\frac{npe_{ehi}}{npecr_{ehi}}$$

Where:

$F_{ehik\ell}$ =   is the weight of the i-th person in the k-th household, from the i-th PSU of the h-th stratum in the e-th entity.

npe_{ehi} =   is the number of selected subjects in the i-th PSU, in the h-th stratum, in the e-th state.

$npecr_{ehi}$ =   is the number of selected subjects in the i-th PSU, in the h-th stratum, in the e-th state.

B) Follow-up sample

The non-response adjustment attributed to the informant, was calculated at domain and entity level using the following expression:

$$F'=F_{ed\ell}^{MHAS}\frac{n_{ed}}{n_{ed}^{*}}$$

Where:

$F_{ed\ell}^{MHAS}$ =   is the MHAS weight of the i-th person, the d-th domain of the e-th entity.

n_{ed} =   is the number of individuals selected for the follow-up in the d-th domain, and the e-th state.

n_{ed}^{*} =   is the number of individuals selected for follow-up, of which an interview was obtained, in the d-th domain and the e-th state.

3.5 Adjustments for projection

The weights of the new sample are adjusted for non-response to ensure a total population for each domain of interest of the survey, given from the projection generated by INEGI, referred to the midpoint of the fieldwork; using the following expression:

$$F_{D}^{''}=F_{D}^{'}\frac{PROY_{D}}{PEXP_{D}}$$

Where:

$F_{D}^{''}$ =   weight adjusted for projection in the domain D.

$F_D^{'}$ = weight adjusted for non-response in the domain D.

$PROY_D$ = population in the domain D, according to the projection.

$PEXP_D$ = total population expanded according to the survey weights, in the domain D.

3.6 Weights

The weight for variable X at the national level is:

$$\widehat{X}=\sum_e \sum_h \sum_i \sum_k \sum_\ell F_{ehik\ell} X_{ehik\ell}$$

The following ratios estimator is used for the estimation of proportions, rates and averages:

$$\widehat{R}=\frac{\widehat{X}}{\widehat{Y}}$$

Where the variable $\widehat{Y}$ is defined in relation to $\widehat{X}$.

3.7 Estimation of sampling errors

The method of the *Last Cluster*[2] was used for the estimation of the sampling errors of the main national estimates, based on the concept that in a multistage design, the largest contribution to the variance of an estimator is presented between the PSU. The term *Last Cluster* is used to denote the total units in the sample of a primary sampling unit.

In addition to the Last Cluster method, the Taylor Series method was used to obtain the precision of the ratio estimators resulting in the following formula to estimate the precision of R:

$$\widehat{V}(\widehat{R})=\frac{1}{\widehat{Y}^2}\sum_{e=1}^{32}\left\{\sum_h^{L_e}\frac{n_{eh}}{n_{eh}-1}\sum_i^{n_{eh}}\left[\left(\widehat{X}_{ehi}-\frac{1}{n_{eh}}\widehat{X}_{eh}\right)-\widehat{R}\left(\widehat{Y}_{ehi}-\frac{1}{n_{eh}}\widehat{Y}_{eh}\right)\right]^2\right\}$$

Where:

$\widehat{X}_{ehi}$ = weighted total of the study variable X in the i-th PSU, in the h-th stratum, in the e-th state.

---

[2] Hansen, M. H. Horwitz, W.N. y Madow, W.G., Sample Survey Methods and Theory, (1953) Vol. 1, pg. 242.

$\widehat{X}_{eh}$ =   weighted total of the study variable X in the h-th stratum in the e-th state.

n_{eh} =   number of PSUs in the h-th stratum, in the e-th state.

L_e =   number of strata in the e-th state.

These definitions are similar to the definitions of the study variable Y.

The precision estimate of the total estimator is calculated with the following expression:

$$\widehat{V}(\widehat{X}) = \sum_{e=1}^{32} \sum_{h=1}^{L_e} \frac{n_{eh}}{n_{eh}-1} \sum_{i=1}^{n_{eh}} \left(\widehat{X}_{ehi} - \frac{1}{n_{eh}} \widehat{X}_{eh} \right)^2$$

The estimates of the standard deviation (SD), design effect (DEFF) and coefficient of variation (CV) was calculated by the following expressions:

$$SD = \sqrt{\widehat{V}(\widehat{\theta})} \qquad\qquad DEFF = \frac{\widehat{V}(\theta)}{\widehat{V}(\theta)_{SRS}} \qquad\qquad C.V. = \frac{\sqrt{\widehat{V}(\widehat{\theta})}}{\widehat{\theta}}$$

Where: $\widehat{\theta}$ =   estimator of the population parameter θ.

$\widehat{V}(\widehat{\theta})_{SRS}$ =   variance estimator of θ under simple random sampling

Finally, the 100(1- α)% confidence interval is calculated as follows:

$$I_{1-\alpha} = \left(\widehat{\theta} - z_{\alpha/2}\sqrt{\widehat{V}(\widehat{\theta})}, \widehat{\theta} + z_{\alpha/2}\sqrt{\widehat{V}(\widehat{\theta})} \right)$$

Distribution of the sample by state   Table 1

Federal State	Total	Sample Follow-up	New sample
National	**20,542**	**14,283**	**6,259**
Aguascalientes	545	415	130
Baja California	483	352	131
Baja California Sur	530	400	130
Campeche	519	389	130
Coahuila de Zaragoza	549	418	131
Colima	534	404	130
Chiapas	555	424	131
Chihuahua	527	396	131
Distrito Federal	545	413	132
Durango	694	563	131
Guanajuato	1,160	800	360
Guerrero	538	407	131
Hidalgo	512	381	131
Jalisco	1,046	679	367
México	559	427	132
Michoacán de Ocampo	885	618	267
Morelos	525	394	131
Nayarit	791	661	130
Nuevo León	562	431	131
Oaxaca	509	378	131
Puebla	1,027	409	618
Querétaro	508	378	130
Quintana Roo	474	344	130
San Luis Potosí	925	392	533
Sinaloa	558	427	131
Sonora	540	409	131
Tabasco	489	359	130
Tamaulipas	516	385	131
Tlaxcala	528	398	130
Veracruz de Ignacio de la Llave	1,120	404	716
Yucatán	518	387	131
Zacatecas	771	641	130

Sample distribution of follow-up and
new sample for biomarkers by state   Table 2

Code	Federal State	Follow-up	New Sample
	Total	**1,874**	**1,346**