The Mexican Health and Aging Study:

MHAS 2012 Data Files Description

Version 2

September 2013

"The MHAS (Mexican Health and Aging Study) is partly sponsored by the National Institutes of Health/National Institute on Aging (grant number NIH R01AG018016). Data files and documentation are public use and available at www.MHASweb.org."

Suggested citation for this document:
MHAS (2013), The Mexican Health and Aging Study: "MHAS 2012 Data Files Description, Version 1, September 2013." Retrieved from www.MHASweb.org on (date).

Table of Contents

I. Introduction

This document describes the data files of the Mexican Health and Aging Study (MHAS/ENASEM) third wave, presenting information on the structure of the files, identifiers, type of interview, and related variables in 2012.

The 2001 baseline survey of the Mexican Health and Aging Study (MHAS/ENASEM) is a national representative survey of individuals born prior to 1951, that is, the population aged 50 or older as of the year 2001. The baseline survey was conducted in the summer of 2001, and a follow-up visit to the same individuals was carried out in the summer of 2003. The sample for the MHAS baseline was selected from residents of both rural and urban areas, from the National Employment Survey (Encuesta Nacional de Empleo, ENE), carried out by the INEGI (Instituto Nacional de Estadística y Geografía) in Mexico. Households with at least one resident of ages 50 or older were eligible to be part of the MHAS baseline sample. If more than one person was age-eligible in the selected households, then one person was selected at random for the study. If the selected MHAS person was married or in a consensual union, with the spouse residing in the same household, then the spouse or partner was also recruited to be part of the MHAS regardless of his/her age.

The 2012 follow-up visit to all panel individuals was conducted in the fall of 2012. In addition, the sample was refreshed by adding a representative sample of the population from the 1952-1961 birth cohorts, as well as their spouses/partners regardless of age. Similar to the baseline interview, the sampling frame for the new cohort sample was the Mexican National Employment and Occupation Survey (ENOE, previously named National Employment Survey, ENE) 2012.

II. Selected and Interviewed

1) Selected Sample

For the 2001 baseline, a sample size with sufficient statistical power for the study objectives was determined to be 11,000 households with persons aged 50 and older in Mexico. The sample was drawn using multistage probabilistic sampling procedures, using the National Employment Survey (ENE) as the sampling frame. A sub-sample of 1,800 households (about 2,550 persons) was randomly selected to obtain anthropometric measures.

For the third wave in 2012, we attempted to interview all panel persons who were alive in 2003. We also attempted to interview those who had responded in 2001 but were not interviewed in 2003. In addition, we added new sample to yield once again a representative sample of the population 50 and older in 2012. Thus, the 2012 targeted sample was composed of two parts as follows:

a) Subjects aged 50 years or older selected in 2001 and their partners –identified at baseline in 2001 or in the follow-up in 2003–, the deceased, those still alive and living in a private dwelling in Mexico (14,283 eligible individuals from the baseline and 385 new spouses).

b) Subjects born in 1952 to 1961 selected in 2012, as well as their spouses/partners that lived with them in a private dwelling in Mexico (6,259 new eligible individuals).

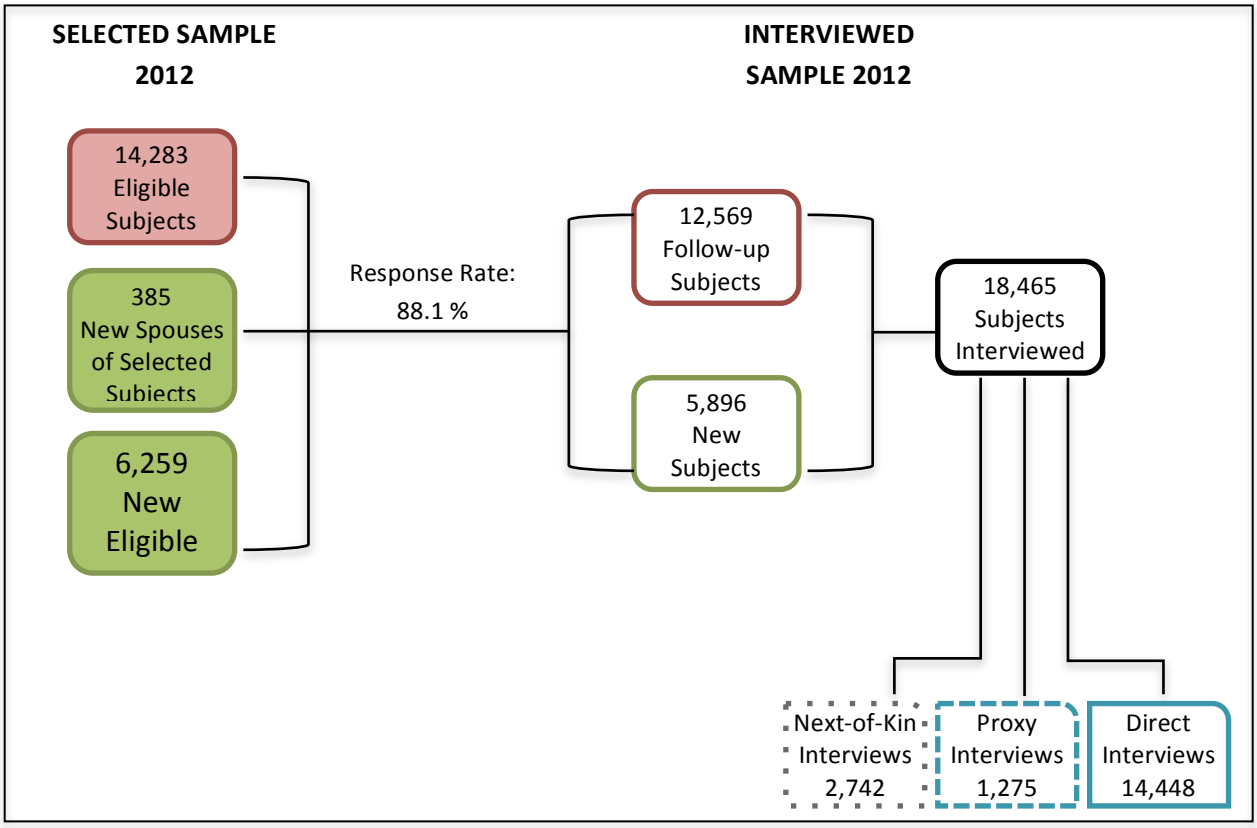
Also similar to the follow-up survey conducted in 2003, the 2012 wave included an interview for the next-of-kin if the study participant had died by the time of the survey.

2) Interviewed sample

In the third wave, the goal was to obtain direct interviews with the study subjects just as described in the previous section. For the follow-up sample, if the study individual was married and the spouse had not been part of the study in the previous wave, or if they remarried between waves, the new spouse was also interviewed regardless of his/her age. Similarly for the new sample, if the selected individual was married or in a consensual union, with the spouse residing in the same dwelling, then the spouse or partner was also interviewed as part of the study regardless of his/her age.

Also similar to the previous waves, if it was not possible to obtain a direct interview due to illness, hospitalization, or temporary absence, a proxy interview was conducted.

In 2012, we completed 18,465 interviews. The figure below provides a summary of the targeted sample sizes as well as the interviewed sample by type of interview conducted.



III. Identifiers

1) 2001 and 2003 Identifiers

Unique Household ID (CUNICAH, also called UNHHID)

In 2001, a random sequential number ranging from 1 to 11,000 was created to identify each household at baseline (CUNICAH also called UNHHID). This variable, in combination with the person identifier CODENT01 (also called PS3) serve as a unique person identifier for the first wave.

Updated Household ID (ACTHOG)

In 2003, a household code was created to capture changes in the situation of the individual or couple interviewed in 2001. This is referred to as “updated household” (ACTHOG) and the

codes reflect the type of change experienced, including divorce/separation, death, or new spouse. In the case of split couples as mentioned above, an interview was sought with both baseline respondents in their respective households and their new spouses if applicable. The updated-household codes also capture whether the household observed in 2003 contains the baseline *sampled* respondent, or the baseline *spouse* of the selected person. Thus in 2003, the unique household identifier CUNICAH used in 2001 is supplemented with ACTHOG to form the unique household identifier. These two variables, in combination with the person identifier for the 2003 given by CODENT03 (also called ENT2) serve as unique identifiers for the second wave.

Unique Person ID (CODENT____)

Within each household, in the initial interview, there are up to two persons under study (the selected person of eligible age, and spouse regardless of age). Once interviewed, the two individuals become part of the longitudinal study.

In the baseline interview, the unique person ID (CODENT01) is assigned in the household as follows (this distinction is important, as it determines the sample weight for each person):

1 = Selected individual,
2 = Spouse.

In 2003, these two values were dragged from 2001 with each respondent, and we added two possible values:

3 = New spouse of the person who was = 1 in 2001, and
4 = New spouse of the person who was = 2 in 2001

In combination with the unique household ID in 2001, and the updated household ID in 2003, these codes uniquely identify individual subjects of study.

2) 2012 New codes of identification

Unique Household ID (CUNICAH also called UNHHID)

In 2012, the random sequential number CUNICAH (also called UNHHID) used to identify each

household at baseline was also assigned to the new refresher sample. The new sample was assigned a value starting from 11,001 to identify each new household.

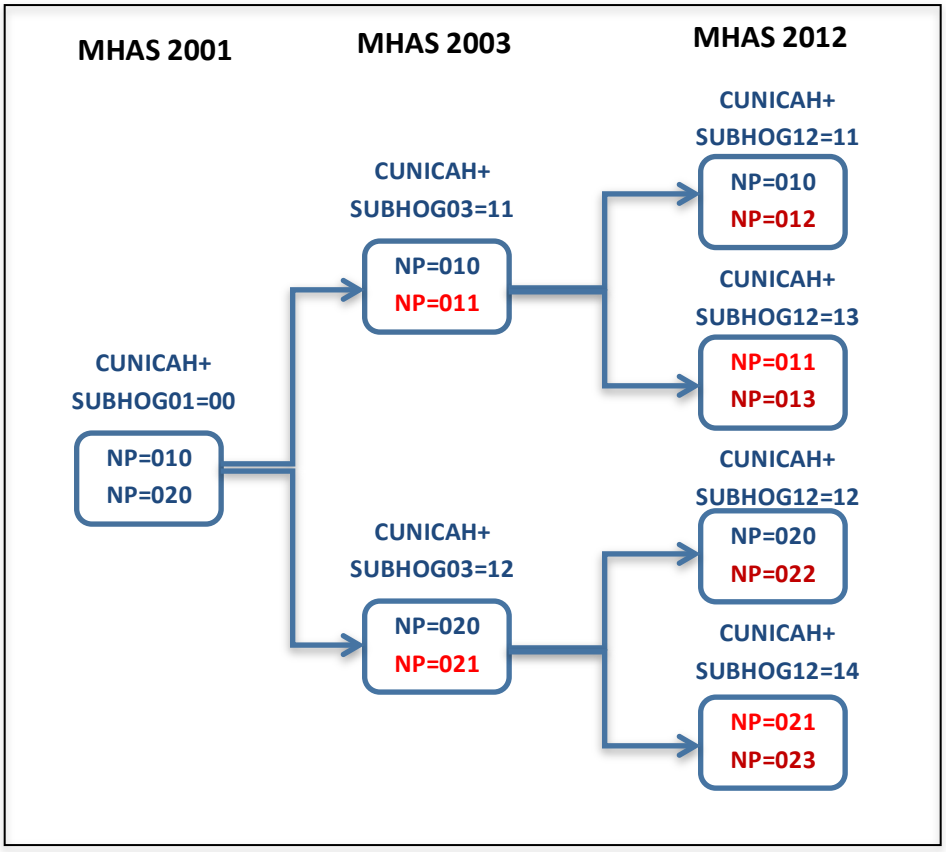
New Person identifier (NP)

Starting in 2012, the variable NP is created to identify each person in the original household, as follows:

NP=10 (selected person), NP =20 (spouse of selected person)

From then on, if these subjects get a new spouse, they will be assigned codes that correspond to the person they are married to (11, 12, etc) if they are spouses of original subject 10; and (21, 22, etc) if they are spouses of original subject 20.

In combination with the unique household ID (CUNICAH), this code uniquely identifies individual subjects of study. The following figure illustrates the possible values of NP and how they are assigned.



Sub-household ID (SUBHOG____)

A sub-household ID (SUBHOGXX) was created to follow the modifications of the original household and new households that result from the changes occurred over time. These changes include: divorce/separation, death, or new spouse of the original subjects. The last two digits of the variable indicate the year of the respective survey. Thus, we constructed one variable for each wave (SUBHOG01, SUBHOG03, SUBHOG12), each of them reflecting the changes in the household recorded for 2001, 2003 and 2012 respectively. This new identifier was created to replace the "updated household" ID (ACTHOG) used in 2003.

At baseline, the variable takes the value 00 to reflect no change in the household (SUBHOG01 = 00) for all households because they are all original households, new to the study in this wave. In the same way, for the new refresher sample of households added in 2012, we assigned the same value (SUBHOG12 = 00) to indicate that these are new to the study in this wave.

In 2012, these two variables in combination – CUNICAH and SUBHOG12 –serve as unique household identifier. The following table indicates all the possible values of SUBHOGXX and the definition for each of them.

VARIABLE: SUBHOGXX	
VALUES	**DEFINITION**
BASELINE HOUSEHOLD	
00	For all households that are new sample (in their baseline year)
NO CHANGES IN THE HOUSEHOLD	
01	The household contains an individual **NP = 010** (first selected subject that entered the sample)
02	The household contains an individual **NP=020** (subject in this household that entered to the sample first)
03	The household contains an individual **NP=011** (subject in this household that entered to the sample first)
04	The household contains an individual **NP=021** (subject in this household that entered to the sample first)
Note: Starting from 05 up to 10 we could assign new IDs in future waves.	
CHANGES IN THE HOUSEHOLD	
Note: The change can indicate a dissolution (due to the death of the spouse or separation – with or without a new union) or a new union. In case of a death in the household, this ID is assigned to the survivor.	
11	The new household contains an individual **NP=010** (subject in this new household that entered to the sample first)
12	The new household contains an individual **NP=020** (subject in this new household that entered to the sample first)
13	The new household contains an individual **NP=011** (subject in this new household that entered to the sample first)
14	The new household contains an individual **NP=021** (subject in this new household that entered to the sample first)
15	The new household contains an individual **NP=012** (subject in this new household that entered to the sample first)
16	The new household contains an individual **NP=022** (subject in this new household that entered to the sample first)
17	The new household contains an individual **NP=013** (subject in this new household that entered to the sample first)
18	The new household contains an individual **NP=023** (subject in this new household that entered to the sample first)

Note: Starting from 19 up to 30 we could assign new IDs in future waves.	
FALLECIMIENTO	
Note: This ID is assigned to the deceased; to the survivor we assigned the ID from CHANGES IN THE HOUSEHOLD.	
31	The deceased is **NP=010**
32	The deceased is **NP=020**
33	The deceased is **NP=011**
34	The deceased is **NP=021**
35	The deceased is **NP=012**
36	The deceased is **NP=022**
37	The deceased is **NP=013**
38	The deceased is **NP=023**
Note: Starting from 39 up to 50 we could assign new IDs in future waves.	
RE-UNION	
70	Two individuals (NP=010 y NP=020) part of the study, they were together, separated, and got back together
71	Two individuals, from the family of 10, they were together, separated, and got back together
72	Two individuals, from the family of 20, they were together, separated, and got back together
NEW INDIVIDUAL	
99	To the new individuals we assign 99 in the previous waves to the one when they enter the study
DECEASED	
88	To the deceased we assign 88 in the posterior waves to the one when he/she was found dead
NOT CONTACTED	
77	To the individuals not contacted (whole household)

IV. Structure of 2012 Files

The MHAS/ENASEM 2012 included three types of questionnaires according to the type of interview: a) direct, b) proxy, or c) next-of-kin, and their status in the study: a) follow-up, for those who completed an interview in the 2001 or 2003, or b) new person. In addition, the household roster, information on children (alive and deceased), and information on assets and income were sought at the household level. Thus the files structure reflects all these types of interviews by individual and household level.

All the variables in the MHAS 2012 data files have a suffix to indicate the year they are associated with and are presented as follows: “01” for 2001 variables, “03” for 2003 variables, and “12” for 2012 variables. Some variables such as the identifiers do not have this suffix.

Below, we give a brief description of the data files by type of interview.

1) Individual Level Files

a) Sections A, C, D, E & PC, F, H, I, and EM (all together in one file)

In this file, we included the sample weights constructed at individual level (FACTORI_12), household level (FACTORH_12), and biomarkers (FACTORA_12). We also included the MHAS 2001 and 2003 identifiers mentioned above (CUNICAH also called UNHHID, CODENT01 also called PS3, CODENT03 also called ENT2) to facilitate the merge of this file with the 2001 and 2003 data files.

Finally, we constructed a series of variables to determine the type and order of interviews, locality size and high-migration state. The variables were created as follows:

Type of interview and status in the study in 2012 (TIPENT_12)

This variable indicates the type of interview (direct, proxy, or next-of-kin), and their status in the study, follow-up or new sample. The following table indicates the possible values and frequencies at the individual level:

Type of interview 2012	Freq.	Percent
Direct, follow-up sample	8,868	48%
Direct, new sample	5,580	30%
Proxy, follow-up sample	959	5%
Proxy, new sample	316	2%
Next-of-Kin	2,742	15%
Total	**18,465**	**100%**

Order of interviews in 2012 (ORDER_12)

This variable indicates the order of the interviews in the household: first, second or third interview. In some cases, we conducted up to three interviews in one household; we found a new union after one of the respondents died between the two waves. In those cases, we conducted a next-of-kin interview, a follow-up interview, and new individual interview.

Sex (SEX_12)

This variable indicates the gender of the study subjects. It was constructed from the sex reported in Section A for follow-up individuals and Section AA for new sample.

Age (AGE_12)

The age variable was constructed by the fieldwork agency that conducted the survey (INEGI) and included the age for follow-up and new respondents. The age reported was validated using the birth date from Section A for follow-up individuals and Section AA for new sample; the result indicated consistency between the age variable and the dates.

Years of Education (YRSCHOOL_12)

The education variable was constructed using the same code used at baseline, only for the new sample. Please visit the Forum using the following link for more details on how to create this variable: http://www.mhasweb.org/DiscussionForum/Default.aspx.

Locality Size (TAM_LOC_12)

A locality size variable was also included. The following table illustrates the variable's possible values and frequencies at the individual level:

Locality size (inhabitants) 2012	Freq.	Percent
100,000+	9,123	58%
15,000-99,999	1,741	11%
2,500-14,999	1,738	11%
<2,500	3,121	20%
Total	**15,723**	**100%**

High Migration States (EAM_12)

A dummy variable that indicates the high migration states that were oversampled, was also included. These are states in Mexico with high proportion of the population that are migrants to

the United States. The variable takes the value 1 if the individual resides in one of these states. The following table shows the distribution of cases at the individual level:

High migration states 2012	Freq.	Percent
0	10,515	67%
1	5,208	33%
Total	**15,723**	**100%**

b) Section Objective Markers (Biomarkers, performance and anthropometric measures)

The MHAS 2012 fieldwork included a collection of intravenous and capillary blood from 2,086 respondents by experienced personnel from INSP (Instituto Nacional de Salud Publica de Mexico). The targeted sub-sample for objective markers was the full sample in four states, including a relatively poor state, a relatively urban state, a high-US-migration state and a high-diabetes state. This selection prioritizes international migration, poverty, and diabetes as research topics. By sampling these 4 states we cover a wide spectrum of urban/rural, age, economic, and social strata, as well as a range of health and disability strata.

This data file includes the MHAS 2001 and 2003 identifiers mentioned above (CUNICAH also called UNHHID, CODENT01 also called PS3, CODENT03 also called ENT2) to facilitate the merge of the 2012 files with the 2001 and 2003 data files, as well as sampling weights for the biomarkers sub-sample (FACTORA_12). The file also includes the following constructed variables: type of interview (TIPENT_12), order of the interview (ORDER_12), locality size (TAM_LOC_12), high migration states dummy (EAM_12), and age and gender (AGE_12 and SEX_12). The variables were created as described above.

c) Next-of-Kin: Sections SA, SB, SC, SD, SE, SH, and SI (together in one file)

The Next-of-Kin data file includes the information from sections SA, SB, SE, SH and SI in one file, which also includes the MHAS 2001 and 2003 identifiers mentioned above (CUNICAH also called UNHHID, CODENT01 also called PS3, CODENT03 also called ENT2) to facilitate the merge of the file with the 2001 and 2003 data files. This file also includes the following

constructed variables: order of the interview (ORDER_12), locality size (TAM_LOC_12), and high migration states dummy (EAM_12) mentioned above.

2) Household Level Files

We constructed a household level data file for each section of the interview that was conducted at the household level:

- TRH (Household roster) – Follow-up sample
- TRH (Household roster) – New Sample
- Section B (Non-resident children) – Follow-up sample
- Section B (Non-resident children) – New sample
- Section B (Deceased children) – Follow-up sample
- Section B (Deceased children) – New sample
- TRH and Section B – This file contains a few variables from the beginning of the sections at the household level
- Sections G, J, and K (in one file)

Each of these household level files includes the sample weight at the household level (FACTORH_12), as well as the MHAS 2001 and 2003 household identifiers mentioned above (CUNICAH from 2001 and 2003, and ACTHOG from 2003) to facilitate the merge of these files with the 2001 and 2003 data files. The files also include the following constructed variables: (TIPENT_12), locality size (TAM_LOC_12), and high migration states dummy (EAM_12), which were described above.

IV. Merging files across waves: 2001, 2003 and 2012 data

All the individual and household level files for 2012 contain the MHAS 2001 and 2003 identifiers as mentioned above (CUNICAH also called UNHHID, CODENT01 also called PS3, CODENT03 also called ENT2) to facilitate the merge of these files with the 2001 and 2003 data file.

The unique household identifier CUNICAH used in 2001 must be used to merge household level files from the first wave. This identifier in combination to the person identifier for the 2001

survey, given by CODENT01 (also called PS3), serve as unique identifiers to merge individual level files.

In 2003, the unique household identifier CUNICAH used in 2001, supplemented with ACTHOG must be used to merge household level files. In addition to these two identifiers, the person identifier for the 2003 survey, given by CODENT03 (also called ENT2), serve as unique identifiers to merge individual level files.

The following figure illustrates the suggested merging process at the individual level.