

---

# The Mexican Family Life Survey

## First Wave

LUIS RUBALCAVA  
GRACIELA TERUEL

PRINCIPAL INVESTIGATORS

February 2007

---

The Mexican Family Life Survey baseline and documentation public release was sponsored by the Ford Foundation. For assistance or questions regarding MxFLS, please contact [mxfls.support@cide.edu](mailto:mxfls.support@cide.edu) or [mxfls.support@uia.mx](mailto:mxfls.support@uia.mx).

Please refer to the following citation for the MxFLS data:

Rubalcava, Luis and Teruel, Graciela (2006). "User's Guide for the Mexican Family Life Survey First Wave", [www.mxfls.uia.mx](http://www.mxfls.uia.mx).

## Preface

This document describes the Mexican Family Life Survey, and the design and implementation of the Mexican Family Life Survey baseline (MxFLS-1). The main goal of the MxFLS project is to field and place in the public domain a rich longitudinal, nationwide, representative database that spans for a period of at least 10 years to better understand the social, economic, demographic and health transitions happening in Mexico and the dynamics of Mexicans who decide to migrate to the United States during the first decade of the 21<sup>st</sup> century.

Fielding of the MxFLS baseline (MxFLS-1) concluded in August 2002, with a sample of 8,440 households. The second wave (MxFLS-2), which will provide the foundation for a long term panel survey, will begin fieldwork in the year 2005 while MxFLS-3 is scheduled to take place in 2008.

MxFLS-1 was a collaborative effort between researchers and officials from Universidad Iberoamericana, AC (UIA), *Centro de Investigación y Docencia Económicas (CIDE)*, the Mexican National Bureau of Statistics (INEGI) and the Mexican National Institute of Perinatology (INPer).

Funding for MxFLS-1 activities was provided by the Mexican Council for Science and Technology (CONACYT), the Mexican Ministry for Development (SEDESOL), the Mexican Social Security Institute (IMSS), the Ford Foundation, the University of California Institute for Mexico and the United States (UC-Mexus) and UIA.

The Second Wave of the Mexican Family Life Survey (MxFLS-2) is a collaborative effort among researchers from UIA, CIDE, the Mexican National Institute of Public Health (INSP), and the California Center for Population Research (CCPR) at the University of California, Los Angeles (UCLA).

The MxFLS public use file documentation are of interest to policymakers concerned about the socioeconomic and health trends of Mexico, to researchers who are considering using the MxFLS data, and to those studying the migration dynamics between Mexico and the United States. Updates regarding the MxFLS database subsequent to publication of these volumes will appear at the MxFLS websites: <http://www.mxfls.uia.mx> and <http://www.mxfls.cide.edu>.

### Documentation for MxFLS-1

1. Rubalcava, Luis and Teruel, Graciela (2006). "The Mexican Family Life Survey First Wave", [www.mxfls.uia.mx](http://www.mxfls.uia.mx).
2. INEGI (2004). "Sample Design". Description of MxFLS Baseline Sample. CIDE & UIA Working Paper.
3. Household Survey Questionnaire for the First Wave of Mexican Family Life Survey. Spanish and English versions.
4. Household Survey Codebooks for the First Wave of the Mexican Family Life Survey. Spanish and English versions.
5. Community Survey Questionnaire for the First Wave of the Mexican Family Life Survey. Spanish and English versions.
6. Community Survey Codebooks for the First Wave of the Mexican Family Life Survey. Spanish and English versions.

### Release of MxFLS-1 Data

The First Wave data of the Mexican Family Life Survey was released on October 22, 2004; [www.mxfls.cide.edu](http://www.mxfls.cide.edu) and [www.mxfls.uia.mx](http://www.mxfls.uia.mx).

## Acknowledgments

We are grateful to Elizabeth Frankenberg and Duncan Thomas of UCLA for their unconditional support to the development of MxFLS-1 throughout all its phases. The design of the questionnaires, methodology and fieldwork protocols of MxFLS-1 draw heavily on the Indonesian Family Life Survey-2 (IFLS-2) and benefited enormously from suggestions made by Frankenberg and Thomas..

We would like to thank INEGI for their collaboration in MxFLS-1. We are grateful to Francisco Javier Gutiérrez for supporting the implementation and development of the survey. We are especially grateful to Patricia Méndez and personnel of the Mexican Income and Expenditure Survey (ENIGH) for the enormous effort put into MxFLS baseline, before, during, and after fieldwork activities were conducted; without her openness and willingness to collaborate the survey would have not been a success.

The survey would not have been possible without the support of authorities from UIA, CIDE INEGI and INPer.

We are grateful to our colleagues Ana Laura Magaloni, Layda Negrete, Susan Parker, Allison Rowland, Juan Manuel Torres and Guillermo Trejo from CIDE; and from UIA to Francisco Calderón, Pablo Cotler, Alba Torres, Joaquín Palomar and Renata Villoro.

We would like to thank Felipe Vadillo who was responsible for the training and supervision of the health personnel that participated in MxFLS-1. Vadillo represented the National Institute of Perinatology and was the researcher responsible for obtaining all health permits from the Ministry of Health in Mexico.

We are indebted to the MxFLS-1 team: Armando Baigts, Itzel Bello, Antonio Bruschetta, Alejandro Celis, Oscar Contreras, Karina Cordova, Miguel Ángel Delgado, Cecil García, Ernesto García, Santiago Guerrero, Ana Lourdes Gómez, Juan Carlos González, Maria José González, Celeste De la Huerta, Iliana Machuca, Fernanda Manzano, Montserrat Manzano, Erika Martínez, Paulina Oliva, Pamela Ortiz, Mónica Palacios, Alan Pasalagua, Vanesa Pérez, José Manuel Prado, Yaeli Reyes, Ingrid Riquelme, Vianey Rosales, Claudia Ruiz, Alicia Santana, Maria Fernanda Santiesteban, Humberto Soto, Amadeo Valdez and Regina Weber.

We are very grateful to MxFLS-1 financial supporters: SEDESOL, CONACYT, Ford Foundation, IMSS, UIA, CIDE and UC MEXUS.

Finally, we are indebted to all our respondents for their patience and interest which made this survey possible.

## **Contents**

Preface	iii
Acknowledgments	iv
<b>1. Introduction</b>	<b>1</b>
1.1 Institutional Collaboration	2
1.2 Project Planning	2
1.3 MxFLS-1 Sponsors	2
<b>2. MxFLS-1 Design</b>	<b>3</b>
2.1 MxFLS-1 Questionnaire Design	3
2.2 Household and Community Information	3
2.3 Confidentiality	4
2.4 MxFLS-1 Household Sampling Design	4
2.5 Household weights	5
<b>3. File Structure and conventions</b>	<b>5</b>
3.1 Files Organization	5
3.1.1 Descriptive documentation	5
3.1.2 Household Survey Files	6
3.1.3 Community Files	6
3.1.4 Weights	7
3.2 SECUENCIA" Variables	7
3.3 Missing Values and Special Codes	7
3.4 Response Types	8
<b>4. Household Survey</b>	<b>9</b>
4.1 Survey Instruments	9
4.2 Data File naming	16
4.3 Variable naming	16
4.4 Identifiers and Levels of Observation	17

4.5 Special Features of the MxFLS-1 Household Data	17
4.5.1 Symmetric Information	17
4.5.2 Duplicate Information	18
4.5.3 Family Relationships	19
<b>5. Community Survey</b>	<b>21</b>
5.1 Survey Instruments	21
5.2 Data File naming	25
5.3 Variable naming	25
5.4 Identifiers and Levels of Observation	25
<b>6. Linking MxFLS-1 Data files</b>	<b>26</b>
6.1 Concatenating Data	27
6.2 One-to-one merges at the Individual, Household, community or facility level	27
6.3 One-to-many Merges	27
6.4 Merging Household data with community or facility data	29
<b>7. Cleaning the MxFLS-1 Data</b>	<b>29</b>
7.1 In the Field: CAFE Editing, Interviewer Rechecks	29
7.2 Double Data Entry and Verification	29
7.3 “Look Ups”	29
7.4 Special Cleaning for Open-ended, “other” and Numeric Variables	30
7.5 Section Checks	31
7.6 Checks on IDs across Books	31
7.7 Checks on Units of Measure	32
7.8 Created Variables and Files	32
<b>8. References</b>	
<b>Appendix A: Response Rates of The Household Survey</b>	<b>33</b>

<b>Appendix B:</b> Names of Data Files for the Household Survey	34
<b>Appendix C:</b> List of public and non public open questions	35
<b>Appendix D:</b> Description of the NAICS (North American Industrial Classification System) and the CMO (Mexican Classification of Occupation).	38
<b>Appendix E:</b> Correct answers for adults and children's Raven test	47
<b>Appendix F:</b> Names of Data Files for the Community Survey	50
<b>Appendix G:</b> Classification of Medicines of book of community prices	53
<b>Appendix H:</b> Linking Of Schools, Clinics And Small Health Providers in Mxfls-1 Community Survey with Mxfls-1 Households	54
<b>Glossary</b>	55

# 1. Introduction

The main goal of the MxFLS is to generate a uniquely rich longitudinal database that spans for a period of at least 10 years to better understand the social, economic, demographic and health transitions happening in Mexico during the first decade of the 21<sup>st</sup> century, as well as the dynamics of Mexicans when/if they decide to migrate to the United States..

Understanding those changes has important implications for the United States and Mexico. Migration between Mexico and the U.S. is a central concern for policy-makers on both sides of the border. Since the signing of the North American Free Trade Agreement (NAFTA), the U.S. has experienced the largest annual rate of immigration from Mexico ever. In parallel remittances from Mexican immigrants living in the U.S. have become the second largest source of capital inflow to Mexico.

A second aim of this project is to provide scientifically sound evidence regarding the effects of migration to the United States on those who move and on those who stay in Mexico. Specifically, MxFLS future waves will track respondents from the baseline that have moved from Mexico to the United States, and interview them. Combining their responses with information gathered from members of their origin household will provide unparalleled opportunities to examine the complex relationships between migrants and their origin household.

A third aim is to provide new opportunities to investigate the effects of behavioral responses to economic upheaval on the well-being of the Mexican population. Several major economic crises have marked Mexico's economic history over the last quarter of the 20<sup>th</sup> century. Understanding how individuals and households have responded to changes, and identifying those that may be vulnerable in the future, calls for a broad-purpose longitudinal survey data such as MxFLS.

A fourth goal of MxFLS is to improve assessments of policy responses to the economic and social changes of the past several decades. For example, Mexico has implemented one of the most ambitious anti-poverty programs ever conceived. In an explicit effort to reduce economic disparities, in 1997, the government created *Oportunidades* (formerly *PROGRESA*). Since 2003 the program has expanded massively and is now being implemented in urban areas. Today, one out of every five Mexicans is enrolled in the program. The long term longitudinal design of MxFLS will provide a vehicle to measure the medium and long-term impacts of the program on a wide array of indicators of well-being and to also track the distributional effects of the program. Studies using the MxFLS datasets will complement the rigorous impact evaluation conducted by the Federal government.

The fieldwork activities of the first wave of the Mexican Family Life Survey (MxFLS-1) were concluded in August 2002. MxFLS-1 is a broad-purpose, multi-topic, nationally representative survey of individuals, households and communities. The baseline covers over 8,440 households in 150 communities across Mexico and is representative of the Mexican population at the national, urban, rural and regional level<sup>1</sup>.

The household survey includes household data and individual-level information for all members living in the household. During fieldwork, one or two key adult household member respondents were asked to provide information about the socioeconomic status and demographic composition of the household. In parallel, one-on-one interviews were conducted to household members age 12 and over. For children younger than 12, fieldwork personnel interviewed a parent or caretaker.

In addition to traditional monetary indicators such as income and expenditures, MxFLS baseline collected information on individual assets in order to describe accumulation decisions over the life course and provide a better measure of lifetime resources. MxFLS-1 also gathered retrospective data on labor decisions, family business and agriculture activities. Historic information on migration and marriage/union decisions for all adult members. Birth histories of women in reproductive age also conform MxFLS-1 questionnaire. MxFLS-1 includes modules on individual time use, health

---

<sup>1</sup> According to President Fox's National Development Plan.



self-assessment, mental health questions and information about in and out-patient health services utilization. The questionnaire includes two innovative crime and victimization sections, at the household and individual level, which were incorporated in order to capture the impact on the well-being of the population of relatively recent and unprecedented nation-wide high crime rates.

All household members were asked to participate in an in-home physical health assessment. The health assessment included the measurement of anthropometry (weight, height, hip and waist circumference), hemoglobin levels and blood pressure. In addition, Raven (cognitive) tests were applied to members 5 to 12 and 12 to 65 years old.

Parallel to the household survey, MxFLS-1 conducted a large community survey operation where public and private schools, health providers, small health practitioners, community leaders and established merchants were interviewed, in order to assess the characteristics of the communities where MxFLS' sampling individuals reside.

MxFLS-1 public release data and documentation can be accessed at no cost from [www.mxfls.cide.edu](http://www.mxfls.cide.edu) and [www.mxfls.uia.edu](http://www.mxfls.uia.edu).

## 1.1 Institutional Collaboration

Researchers from the *Centro de Investigación y Docencia Económicas (CIDE)* and *Universidad Iberoamericana (UIA)* joined efforts to maximize the probability to conclude satisfactorily the first wave of the MxFLS-1. Fourteen researchers from the two institutions actively participated in the design of the MxFLS-1. Active technical support was received by the MxFLS-1 Scientific Committee, which was integrated by international scholars that have extensive experience in the design and fieldwork of similar surveys in developing countries. Special collaborative mechanisms were established with faculty members of the Department of Economics and Sociology of the University of California Los Angeles (UCLA).

To maximize the success of MxFLS-1 fieldwork, the MxFLS team held additional agreements with two other Mexican Institutions. The National Institute of Perinatology (*INPER*) which was in charge of training the health personnel and collecting anthropometry and biomarkers of MxFLS-1' household members. The Mexican National Bureau of Statistics (*INEGI*) provided the sample design, carried out the socio-demographic fieldwork and coordinated the pre and post-fieldwork activities.

The systematization of the Mexican Family Life Survey baseline data and the generation of public release documentation was sponsored by the Ford Foundation and was mainly carried out by researchers from UIA and CIDE.

## 1.2 Project Planning

Both the planning and direction of MxFLS-1 are responsibility of Luis Rubalcava and Graciela Teruel.

## 1.3 MxFLS-1 Sponsors

Pre-fieldwork activities, fieldwork data collection, and public use information of MxFLS-1 have been supported by several institutions. Their financial contribution was decisive for the completion of the MxFLS baseline. MxFLS-1 team is grateful to the following institutions:

- Mexican Council for Science and Technology [CONACYT]
- The Ford Foundation
- Mexican Ministry of Social Development, [SEDESOL]
- Mexican Social Security Institute, [IMSS]
- The University of California Institute for Mexico and the United States, [UC MEXUS]
- Universidad Iberoamericana, [UIA, FICSAC]

## 2. MxFLS-1 Design

### 2.1 MxFLS-1 Questionnaire Design

MxFLS-1 questionnaires follow the design of the *Indonesian Family Life Survey (IFLS)*, after carefully being adapted to the Mexican context. In the design of MxFLS instrument, MxFLS team took into consideration the design of other Mexican household national representative surveys to maintain comparability, such as the National Income and Expenditure Survey (ENIGH), the National Employment Survey (ENE), and the National Survey on Demographic Dynamics (ENADID).

During the planning stage, MxFLS team coordinated the design of the questionnaire and the pre-fieldwork activities of MxFLS-1. Ten inter-institutional seminars were organized to discuss the content of MxFLS and different drafts of the instrument. A significant number of pilot tests were conducted to test the consistency and the flow of the questionnaire. The pilot tests were applied to different social strata, in urban and rural communities, as well as in indigenous communities.

Once this stage was concluded, the sections were integrated into a unique questionnaire and two field pretests were conducted in order to refine the instrument in its totality. The second pretest—of larger dimensions than those of the first one—involved the participation of INEGI's staff and INPer's health personnel, allowing a complete simulation of the fieldwork operation.

MxFLS-1 training was conducted at INEGI's headquarters in the city of Aguascalientes during ten weeks. Coordinators, supervisors, health workers, household interviewers, community interviewers and editors took the training. MxFLS team participated actively in this process through the direct training of the instrument contents and fieldwork protocols. To achieve the highest possible degree of standardization, the MxFLS-1 team in collaboration with INEGI and INPer personnel elaborated detailed manuals of both the household and the community questionnaires, as well as of fieldwork procedures.

### 2.2 Household and Community Information

MxFLS-1 collects—in one single database—detailed information about households and communities.

Household data includes:

- Expenditure and auto-consumption at household level.
- Saving decisions; assets ownership (wealth), debt; labor and non-labor income; public and private transfers; and participation in social programs at the household and individual levels.
- One year retrospective individual information on formal and informal credits and loans.
- Schooling, and retrospective information about school attendance and grade repetition of every household member.
- Two-year retrospective information about labor decisions of every household member, including children.
- Leisure and time use
- Demographic information about extended non co-resident family of every adult household member.
- Lifetime information about permanent migration (change of residence) and two-year retrospective information about temporary (circular) migration.
- Individual self reports about general health status, chronic diseases, morbidities health related habits and ADLs (for household members above 50).

- A module of mental health comprised of 21 questions.
- Anthropometry and biomarkers for every household member (Height, weight, hip and waist circumference, blood pressure, and hemoglobin level).
- Detailed retrospective information on inpatient and outpatient (health) services utilization.
- Pregnancy histories, contraception use, as well as pre and postnatal services utilization, for women in reproductive age (14-49).
- Marriage and union history of every adult female and male in the household.
- Individual and household crime incidence history (robbery, street crime, etc).
- Detailed information on family business and land use.

Parallel to the household interviews, MxFLS-1 conducted fieldwork activities at the community level. MxFLS-1 baseline included the application of a community questionnaire, with the objective of embodying qualitative and quantitative information at community level on schools and health providers; local prices of goods and services; and community general socioeconomic characteristics.

The community questionnaire includes retrospective information on 1,172 schools, 904 health centers, 1,167 private small health providers, 419 stands, stores, supermarkets, farmers' and established markets and on 166 drugstores. MxFLS-1 community fieldwork conducted interviews to school, health facilities personnel as well as to private health practitioners. Information on local prices and availability of several goods and services were gathered through visits to markets, supermarkets and small stores in the community. Information about community infrastructure and community socioeconomic characteristics was collected through interviews with local community leaders.

## 2.3 Confidentiality

To assure the confidentiality of MxFLS' respondents, all information that could be used to identify respondents has been suppressed from the public use data files. This includes the respondents' names and addresses; the names of the communities they have lived in the past, as well as the names and addresses of the schools, health facilities and health private practitioners that the household members mentioned during the household survey and which later became part of the community survey. In order to assure the confidentiality of the local facilities of the community survey, MxFLS team replaced INEGI's community sampling coding with an internal survey generated code in the public release data files.

## 2.4 MxFLS-1 Sampling Design

INEGI provided the sampling design of MxFLS baseline. The study's universe is shaped by the private dwellings that existed in Mexico in the year 2002. Primary sampling units were selected considering that representation at the national, urban-rural<sup>2</sup> and regional levels was the objective. Regional definitions are in accordance with the Mexican Federal Government National Development Plan 2000-2006. MxFLS-1 shares the same sampling framework that the 2002 Mexican National Employment Survey (ENE-2002).

The sampling design is probabilistic, stratified, multi-staged and by cluster. It is probabilistic because all the sampling units have a known (and different from zero) probability of being selected. It is stratified because sampling units with similar geographic and socioeconomic characteristics were grouped into strata. It is multi-staged because the last unit of selection was chosen after

---

<sup>2</sup> An urban community is defined as a community with a population above 2,500 inhabitants.

several stages; and by cluster because the units of selection are groups of sampling units. The sampling selection of MXFLS-1 is independent at each dominion of study. The sample size is 8,440 households with 35,677 individual interviews.

By design, MxFLS-1 has oversampled rural communities with less than 2,500 inhabitants. Therefore, we recommend to use the appropriated weights when estimating national parameters. (See section on Weights).

For further description about sampling design, please refer to MxFLS-1 sample desing documentation

## **2.5 Weights**

Weights are provided for the household survey data.

MxFLS weights take into account the non-response rate of each book independently. Therefore, each book has a different set of weights. Individually-based books (3A, 3B, 4, 5, EA, EN and biomarkers) have and expansion weight at an individual level, whereas household-based books (C, I, II) have weights at a household level. There are no expansion weights for the proxy book.

To use the expansion weights one needs to merge the information subject to analysis, with the corresponding file containing the related weights. For household-level books the weight file should be merge using the variable "folio." For individual books one should use the variables "folio" and "Is". The variable "factor\_#" assigns the correct weight to each observation.

## **3. File Structure and Conventions**

This section describes MxFLS-1 data and documentation files placed in the public domain. We recommend the use of MxFLS-1 data files in parallel consultation with MxFLS-1 user's guide, survey questionnaires and variables codebooks.

### **3.1 File Organization**

The MxFLS-1 files are organized in three main groups:

- Descriptive documentation of the First Wave of Mexican Family Life Survey
- Household Survey files
- Community Survey files

#### **3.1. 1 Descriptive documentation of the First Wave of the Mexican Family Life Survey**

Description and Documentation Files includes:

- Rubalcava, Luis and Teruel, Graciela (2007). "The Mexican Family Life Survey First Wave", [www.mxfls.uia.mx](http://www.mxfls.uia.mx).
- INEGI (2004). "Sample Design". Description of MxFLS Baseline Sample. CIDE & UIA Working Paper.

#### **3.1. 2 Household Survey Files**

Household survey files include data files, questionnaires and codebooks. Files containing data are available in Stata 8.0 format. For users that would like to work with another file format, we recommend the use of the software Stat/Transfer. Stat/Transfer converts Stata files to other formats such as ASCII, dBase, Excel, SAS, SPSS, Statistica and SYSTAT.

##### **Household Data Files**

- One zipped file containing the ten individual household survey data files that constitute the household survey. *hh* stand for household; *02* stands for the year 2002 of the baseline survey; and *dta* indicates the file contains data in STATA format.
- Ten individual household survey data files (hh02dta\_byy), one for each module (book) of the household survey questionnaire; where *dta* stand for data and *yy* for the equivalent book label (1, 2, 3a, 3b, 4, 5, s, ea, en and x for proxy book).

For every data file of the household survey, there is a corresponding questionnaire file and a codebook file:

### **Household Questionnaire and Codebook Files**

- Ten individual questionnaire files, one for each book of the household survey questionnaire (hh02q\_byy); where *q* indicates questionnaire.
- Ten individual codebook files, one for each book of the household survey questionnaire (hh02cb\_byy); where *cb* stands for codebook
- One zipped file containing the ten individual questionnaire files (hh02q\_all).
- One zipped file containing the ten individual codebook files (hh02cb\_all).

The questionnaire files and codebook files are in PDF format.

### **3.1. 3 Community files**

MxFLS-1 community survey public files, like the household survey files, include data set files, questionnaires and codebooks files. See section 5 for a description of the community survey. The community survey public files display the same structure of the household survey files:

#### **Community Data Files**

- One zipped locality or community survey data file (loc02dta\_all), containing all data set files that constitute the community survey.
- Five separate community survey data set files (loc02dta\_byy), each for every book of the community questionnaire; where *yy* stands for the corresponding book label: Book of Community Characteristics (cc), Book of locality Prices (p), Book of Schools (sch), Book of Health Service Infrastructure (hi), and Book of Small Health Providers (hp).

#### **Community Questionnaire and Codebook Files**

- Five individual questionnaire files, one for each book of the locality or community survey questionnaire (loc02q\_byy); where *q* indicates questionnaire.
- Five individual codebook files, one for each book of the community survey questionnaire (loc02cb\_byy); where *cb* stands for codebook
- One zipped file containing all five questionnaire files of the community survey (loc02q\_all).
- One zipped file containing all five codebook files of the community survey (loc02cb\_all).

The questionnaire files and codebook files are in PDF format.

Users who are only interested in specific sections of the household survey may find convenient to only download the appropriate individual household survey data file, and its corresponding questionnaire and codebook, and weight data files.

MxFLS community survey sample is not probabilistic. The sample was drawn according to what respondents answered about health and educational services which they either attended or knew about. All information from all respondents on each of these services was pulled together to make the selection for the community sample. Those most mentioned were survey with probability one. Those less mentioned still had a positive probability of being selected. In the case of schools, due to the timing of the year when the survey was conducted, all schools mentioned were surveyed. There are no weight files for this section of the survey..

All data and documentation public files of MxFLS have been compress into zip files for easy downloading. In order to decompress each zip files it is necessary that MxFLS users have the software WinZip installed in their computer. WinZip can be freely downloaded from <http://www.winzip.com/downwz.htm>.

### **3.1.4 Household Weight Files**

There is an equivalent weight file for each household survey data file.

- Nine individual weight data files, one for each book of the household survey questionnaire, excluding the proxy book (hh02w\_byy); where w stands for weight file and yy for the corresponding book of the household survey questionnaire.
- One zip file containing the nine individual weight data files (hh02w\_all).

Users who are only interested in specific sections of the household survey may find convenient to only download the appropriate individual household survey data file, and its corresponding questionnaire and codebook, and weight data files

### **3.2 “SECUENCIA” Variables**

MxFLS retrospective data is arranged in grids. Each grid corresponds to a particular event and provides the information of that event. There are as many grids as the number of individual events that constitute a particular retrospective information. The sequence of grids in any retrospective information is identified by the variable name SECUENCIA. In these cases there are multiple records per household or individual. Combining the variables FOLIO and LS with SECUENCIA uniquely identifies an observation. SECUENCIA data is always numeric.

### **3.3 Missing Values and Special Codes**

For numeric variables, a dot (.) signifies missing data. For string variables, a blank space signifies missing data. Variables with missing information may be the result of data properly absent because of skip patterns in the questionnaire. For example, question ah01 in book 3A asks the interviewer to check whether the respondent already answered book 2, and if so, to skip to the next section. If the answer is Yes (ah01=1) the entire section AH will be filled with dots or blank spaces. Sometimes valid answers have no data because the respondent refused to answer the question, or did not know the answer. In these cases special codes ending in 5, 7, 8, or 9 were used.

These special codes can take the following values depending the space provided in the questionnaire for each question:

5, 95, 995 or 9995 = out of range, answer does not fit available space  
7, 97, 997 or 9997 = respondent refused to answer  
8, 98, 998 or 9998 = respondent did not know the answer  
9, 99, 999 or 9999 = question is not applicable

### 3.4 Response Types

The majority of MxFLS-1 questions required either a numeric value or a closed-ended categorical value. However, there are cases where an open-ended response was allowed.

The numeric questions specified the maximum number of digit places allowed in an answer; when any response did not fit, a special code was used by the interviewer, and subsequently the special codes were reviewed and recoded later.

Typically closed-ended categorical questions take dichotomous values for an answer. For example, "Was the school you attended public or private?" When only one answer was allowed, numeric response codes were specified. When more than four numeric response codes were possible, two digits were used so that 95-99 could serve as special codes.

Some questions allow multiple answers, as well. For example, "Who in your family decides about your children's education?" In this case, alphabetic response codes were specified. One different variable was created for each option so that the new variables contain 1 if the respondent answered affirmatively and missing value when answered negatively.

For categorical variables, the questionnaire provides the full meanings for each response category. The codebook contains a brief labeling that summarizes the response category. The questionnaire provides the clearest explanation of each response categories and users should not solely rely on the codebook information.

The codebooks also provide the the distribution of responses. For categorical variables the frequency distribution is provided. For categorical variables where multiple responses were allowed, the codebook provides the number of respondents who gave each response. Since many combinations of responses were possible, the codebook does not provide the distribution of all responses.

## 4. Household Survey

### 4.1 Survey's structure

The household survey includes household data and individual-level information for all members living in the household. During the fieldwork, one or two key adult household member respondents were asked to provide information about the socioeconomic status and demographic composition of the household. In parallel, one-on-one interviews were conducted to household members age 12 and over. For children younger than 12, fieldwork personnel interviewed a parent or caretaker. In addition, a proxy book was fielded to collect information about household members that could not be interviewed in person, mainly because they were absent during the time of the fieldwork. The proxy book includes thematically sections of the questionnaire designed for individual one-on-one interviews. This information comes in limited format and was provided by a household members knowledgeable about the absent member. In a small number of cases, proxy books were also used to complete inconclusive information of one-on-one individual interviews.

For the sake of counting with the most complete information of the household and its members, we recommend to use household level, individual one-on-one interview, and individual proxy information simultaneously.

The household survey questionnaire is divided into seven books which are addressed to different respondents. Each book is in turn divided into topical sections. Books C, 1, and 2 collect the household level information. Books 3A, 3B and 4 collect the information of the adult household members (15 years of age and older). Book 4 is only applied to female respondents in reproductive age (14 to 49 years old). Book 5 collects information of household members younger than 15. Finally, there is the proxy book.

We now describe the content of each book from the household survey questionnaire.

## Books that collect household level information

### **Book C: Control Book and Household Roster.**

Book C serves as a household roster. It is the starting point of the household interview, and it is usually answered by the household head, the spouse or by an adult member knowledgeable about the household. Book C gathers basic data about the demographic and economic characteristics of the members who integrate the household, such as their name, gender, age, schooling, family kinship within the household, whether the individual is absent from the household, and about the individual's income. Book C serves as a control book. The demographics of the household dictate the distribution of books in the household survey questionnaire to be answered during the course of the interview.

Book C also records the geographical information of the household, whether the household was found and if the household interview was conducted. It also gathers information about the characteristics of the dwelling and about the public services available to the household. Book C also collects information on the household plans to change residence. This information is helpful in planning for subsequent rounds of data collection and in tracking respondents. The names of the household members, the household address and any information about moving plans have been suppressed in the public-use data to protect respondent's confidentiality.

We next list the sections that integrate Book C.

**Definitions.** Section DF explains the criteria for basic definitions such as household and household members.

**Household roster.** Section LS collects information about every household member. It gathers basic information on age, sex, marital status, relationship to the head of the household, presence in the household of the individual's mother, father, and spouse, whether the respondent works or is in school, and each household member schooling level.

**Dwelling characteristics.** Section CV asks the respondent to provide information about the characteristics of the dwelling. Section CVO contains the observations from the interviewer regarding the dwelling and its sanitation.

**Information for recontact.** Section RC includes information about the name and address of a local family or friend who might be able to provide location information in the future if the household decides to move. This information has been suppressed from the public use data.

**Family planning and health services.** Section SP collects detailed information about the location of the clinics, hospitals and small health providers that each household reported going or knowing about. This information is not publicly available.

**School infrastructure.** Section EH collects detailed information about the location of primary, secondary and high schools that any household member reported attending. This information is not in the public use data.

### ***Book 1: Household Expenditures***

This book records information about household expenditures and about quantities and purchase prices of several staples. It is typically answered by a female respondent, either by the spouse of the household head or by an adult member knowledgeable about household expenditure affairs.

**Consumption.** Sections CS records detailed information on expenditures for a variety of food and nonfood goods and services, including self-production or in-kind transfers in the last week; personal



care and household items brought during the last month and last three months; and durable goods bought in the last year. For the most relevant products in the Mexican diet (corn tortillas, bakery or store unpacked bread, chicken escalope, steak, pasteurized milk, hen eggs, red tomatoes, beans packed or in bulk, white sugar and soft drinks) information about purchased quantities was collected in addition to expenditure values. Market prices for these standardized products were also gathered through direct interviews to economic establishments in the community (see section 5.1).

### **Book 2: Household Economy.**

This book is usually answered by the household head or the head's spouse or by an adult member knowledgeable of the household economy. Book 2 collects information about household businesses and agricultural activities, household assets and household non-labor income which includes private and public transfers in addition to public program participation.. Other sections of Book 2 collect information about economic and demographic shocks experienced by the household in the previous five years. Book 2 also provides information about retrospective crime events in terms of robberies and assaults that took place in the dwelling or in other household owned properties.:

**Family farm and non-farm businesses.** Sections SU, INR and NNA are focused on household revenues, expenses, and value of assets, household-owned agricultural and non agricultural businesses. SU asks explicitly about land ownership, regardless whether the household farms any land itself. Sections INR and NNA are organized around each family enterprise, differentiating from agricultural and non agricultural businesses.

**Household assets.** Section AH provides data on household assets and assets ownership. Book 3A provides information on assets at the individual level for coupled or married household members..

**Household credit.** Section CRH collects information about the household credit, and on the institutions that provided the loan during the last twelve months.

**Household non-labor income.** Section IN provides data on household non-labor income by its source. This includes income from government supporting programs such as *Progresa* (currently known as *Oportunidades*), *Procampo*, *VIVAH*, *Crédito a la Palabra*, *Alianza para el campo*;<sup>3</sup> education scholarships, special payments received by the household, pensions received by any member of the household, income from asset sales, and returns to assets.

**Economic shocks.** Section SE displays data on economic and demographic shocks experienced at the household level for the past five years such as death or illness of any household member and economic losses due to natural disasters.

**Victimization and Crime history.** Sections VLH and VLH1 contain retrospective information on robberies suffered at the household level and other household own properties.

### **Book 3A: Adult Information (part one)**

Book 3A was applied to all household members 15 years and older. It collects two-year retrospective information about labor decisions; and histories about schooling, marriage/union and migration decisions. Book 3A also includes questions on decision-making and asset ownership for couples.

**Education history.** Section ED displays data on highest schooling level attained. Detailed schooling information is provided for respondents younger than 25 years. The information includes the name, location, and type of school, and grade repetition. Details about schooling expenses,

---

<sup>3</sup> Please refer to the glossary section for a basic description of the government programs listed in MxFLS-1 questionnaire.

classroom size, travel time to school, and if the respondent worked while attending school is in addition collected for individuals who attended school during the past two years. The data about the name and location of the school is suppressed from MxFLS-1 public data.

**Schooling interruptions.** Sections IE and IE1 explores schooling interruptions of household members who suspended their studies for more than 4 weeks in the last five years. The sections give specific information about every interruption mentioned by the respondent.

**Individual non-labor income and assets.** Section AH provides data on individual assets for household members living in couple. Section IIN provides information on non-labor income by source.

**Marriage history.** Sections HM and HM1 displays a complete marriage history including characteristics of the former or non-resident spouses.

**Household decision-making.** Section DH asks couple household members how they collectively make decisions on issues like food preparation, education and health care for children, transfers to extended family members as well as on, labor decisions and use of contraceptives methods.

**Permanent migration history.** Sections MG collects migration histories for every change of residence where the responded lived for a period of 12 months or longer. Data is provided about the respondent's birth place the place of residence at 12 years old, and each subsequent location where the individual lived for a period of 12 months or longer. Information about the individual's migration sequential movements is provided at the state level to protect respondent's confidentiality.

**Temporal or circular migration .** Sections MT collects two-year retrospective data on moves or travel decisions that lasted for more than one month but less than a year. For each move, data were collected on dates and locations, motivations for moving, and distance moved.

**Employment retrospective information.** Section TB provides in depth data about respondents' two-year retrospective labor decisions. Work was defined broadly to include formal and informal, full-time, part-time and seasonal jobs. Data on job occupation and industry sector, type of employer, hours of work and earning for workers and self employed were recorded for primary a secondary jobs. Open ended questions on occupation have been classified according to CMO codes for occupation and NACIS (SCIAN in Spanish) codes for activity, respectively.. (See Appendix C).

**Adult time allocation.** Section ATA asks respondents for leisure use.

**Individual shocks.** Section SHI gathers information about individual perceptions of their life course in the last 12 months and expectations for the next year.

**Crime and victimization history.** Sections VLI provides lifetime retrospective information on street crime events suffered by the respondent such as robberies, assaults, sexual abuse and kidnapping.

### ***Book 3B: Adult Information (part 2)***

Book 3B collects information about health related habits, self reported health conditions, health insurance, inpatient and outpatient health services utilization, and information on mental health to measure the prevalence of anxiety and depression. Book 3B also provides data about no co-resident extended family members (parents, siblings, and children). Data is recorded about the flow of monetary and in-kind transfers to and from the extended family members and other no co-resident individuals.. This section of the questionnaire also provides data on individual credit credit decisions during the last twelve months.

**Tastes and habits.** Section GH provides data on health related habits such as smocking, drinking and exercise.

**Health status.** Section ES collects information about the individual's perception of his/her general health status; about any serious health conditions the person may have phased during lifetime; and information about four week recall morbidities. Section EC provides data on selfreported chronic diseases, and any health expenditure related to these chronic conditions,

**Mental Health.** Section SM asked about individual's own perceptions on emotional aspects of their lives. It draws from mental health questions tested and validated by the National Psychiatric Institute in Mexico (Caraveo 1998).

**Health services utilization.** Section CE collects detailed information on outpatient visits during the last four week. Section HS provides detailed data on inpatient visits (hospitalization) during the previous 12 months. Respondents were also asked about the type and cost of self-medication during the past four weeks (section ATS).

**Health insurance and health care.** Section CA collects data on any health insurance and health care benefits that the individual may have.

**Recontact information in the United States.** Section RE collects information whether the individual has a relative living in the United States. The sections contains information on the address, phone number of any relative. This information is not for public use.

**Credit.** This section provides data on the respondents' knowledge about credit institutions (and lenders) to borrow money and whether the respondent has tried to borrow any loan during the past 12 months. For every attempt to borrow a loan the section provides detailed information the loan, such as the period and scheme of repayment and the collateral requirements, among others.

**Non-resident extended family.** Sections TP, TH and THI collect demographic characteristics and place of residence about the respondent's parents, sibling and children that do not reside in the household. The information is provided along with complete data on the net transfer flow (monetary and in-kind) in the past 12 months between the respondent's and the extended family. Section TO additionally collects net transfers information between the respondents and other non-resident individuals such as other family members (besides parents, sibling or children), non-resident spouse and friends or neighbors.

#### ***Book 4: Information about woman***

Book 4 was administered to all female household members between 14 and 49 years. Book 4 provides data on pregnancy and birth outcome histories; retrospective information on prenatal, postnatal care and nursing practices for every pregnancy and birth outcome in the last five years. Book 4 also provides information on contraceptive use.

**Pregnancy summary.** Section RES provides a summary of all pregnancies.

**Pregnancy history.** Section HE provides historic information about pregnancy outcomes. For all live births data is provided on the child's gender, size and weight at birth and breastfeeding practices. In addition, for pregnancies in the last five years, respondents were asked about prenatal and postnatal care utilization services (number of care visits and type of services received, length of labor, place of child birth). MxFLS public release data allows to merge each child household member with his/her corresponding pregnancy data if the mother answered book 4. To do so each live birth in book 4 carries the corresponding child household member (Is) identification number. Live births with no corresponding child household member have been coded with '99'.

**Knowledge and use of contraceptive methods.** Knowledge, use and market prices of contraceptive methods is directly provided by the respondent in section AC.

#### ***Book 5: Child Information.***

Book 5 of the questionnaire collects information about children younger than 15. For children younger than 12, the child's mother, guardian, or caretaker answered the questions. Children between the ages of 12 and 15 years old were allowed to respond for themselves if they felt confident. Book 5 parallels different sections of books 3a and 3b with the appropriate age-range modifications. Book 5 provides information about the child's schooling, general health status, morbidities, acute health conditions, health services utilization and leisure and labor decisions

### **Antropometry and biomarkers**

MxFLS-1 collected physical health assessments on every household member in addition to the health condition data of books 3B and 5.

#### *Physical Health Assessments*

In addition to collecting data on the respondent's perception about his or her own health status, the MxFLS protocol entailed conducting physical health assessments during the fieldwork. Health workers collected antropometric measures and biomarkers to every household member during the interview. They comprise measures on height, weight, waist and hip circumference, resting blood pressure and hemoglobin levels. Hemoglobin was measured by blood from a finger stick analyzed by a Hemocue photometer brought to the respondent's home. In addition, prior to conducting the physical health assessments, each respondent was asked to provide his or her own perception about his/her height and weight.

### **Books EA and EN: adult and child cognitive ability**

During the Survey interview, household members between 5 and 65 years old were asked to take a test of Raven's Progressive Matrices. The Raven tests are designed to measure the person's cognitive ability and do not require that the person is literate. To avoid possible distractions during the test, household members were asked to take the test alone and in a quiet place within the dwelling. Book EA contains the Raven's 12-colored progressive matrices test applied to children in the age of 5 to 12 years old. Book EN contains the Raven's 18-progressive matrices test applied to adult household members between 13 and 65 years old. Appendix E provides the solution of the tests. For more information about Raven's test please refer to:

- Raven, J.C., Court, J.H., & Raven, J. (1986). Manual for Raven's Progressive Matrices and Vocabulary Scales (Section 2) - Coloured Progressive Matrices (1986 edition with U.S. norms). London: Lewis.
- Raven, J.C., Court, J.H., & Raven, J. (1983). Manual for Raven's Progressive Matrices and Vocabulary Scales (Section 3) - Standard Progressive Matrices (1983 edition). London: Lewis.

### ***Book Proxy: Adult Information by Proxy.***

MxFLS fieldwork protocol was designed to interview every household member via one-on-one interviews. Nevertheless some household members could not be interviewed in person either because they were absent during the time of the fieldwork, or because they suffered from a health condition. In these cases a proxy respondent was asked to provide the information of the missing member. The proxy book contains shortened versions of most of the sections included in books 3A, 3B, and 4 of the questionnaire and was answered by a household member knowledgeable about the absent member. In some cases, proxy books were also used to complete inconclusive

information of one-on-one individual interviews. *It is important to mention that proxy books were only used as a last resource to collect data of a household member.*

### Differences in Information collected from Proxy Book vs. Individual Books

Section	Information in proxy Book	Additional information in Main Book
HM - Marital history	Information about current/last spouse/couple.	History of marriages (book 3A)
MG - Permanent migration: one year or more	Residence information at birth, at 12 years old and current location	History of migrations (book 3A)
ED - Education	Literacy, highest schooling level and schooling expenditures.	Detailed schooling information for each formal schooling attendance level elementary, secondary, high school and university). (book 3A)
TB - Employment	Information about current labor decisions, including occupation earnings and hours worked for primary and secondary job.	Detailed information on current and two-year retrospective labor decisions (book 3A)
CR – Credit	Data on credit and loans during the last 12 months.	No differences (book 3B)
GH - Tastes and habits	Drinking, smoking and exercise habits	No differences (book 3B)
ES - Health condition	Data on general health status, morbidity, chronic diseases and silent health conditions during in lifetime.	No differences (book 3B)
CE - Outpatient Utilization	Data on outpatient health services utilization during the last four weeks.	Detailed information about outpatient health services utilization during the last four weeks.. (book 3B)
HS - Inpatient Utilization	Inpatient health services utilization	Detailed data on inpatient health services utilization (book 3B)
CA - Insurance condition	Data on health care and health insurance.	No differences (book 3B)
TP - Non-resident parents transfers	General information on non-resident parents, including transfers.	No differences (book 3B)

TH - Non-resident siblings transfers	General information on non-resident siblings, including transfers.	No differences (book 3B)
THI - Non-resident child transfers	General information on non-resident children, including transfers.	No differences (book 3B)
TO -Transfers of other non-resident persons	Information on transfers related to other non-resident extended family members, friends and neighbors,	No differences (book 3B)
RES - Pregnancy summary	Women respondents older than 14 and younger than 49 provided information about all pregnancies	No differences (book 4)
HE - Pregnancy history	Detailed information about the last two pregnancies such as name and address of place where attended, child's health after birth and feeding practices.	Detailed information about lifetime pregnancies and births. (book 4)
AC- Contraception	Current contraceptive methods used by the respondent.	Whether they had ever heard of a number of modern and traditional contraceptive methods, whether they had ever used each method, and, if appropriate, whether they knew the price and where to obtain the method. (book 4)

### Supplements.

Sections of MxFLS questionnaire that collect repeated information for different observations are arranged in grids. Each grid provides detailed information of an individual observation, and within each section of the questionnaire, there are as many grids as number of observations or the interviewer reports. Repeated observations can be related to current, retrospective or historic events. For example, , section SU of Book 2 collects detailed information on agricultural production about every plot or parcel the household owns.. Because households may have more than one plot, the same questions about the size, the state of ownership and the use of the plots are asked along grids for every parcel.. Another example is section LS of Book C, which contains the household roster, and records the name, age, birth date and sex for each household member.

The printed questionnaire provides only space to record data for up to five different plots and up to ten household members in sections SU and LS, respectively. However, additional supplements were used in the event that the interviewer listed more observations or events than those that fit in the questionnaire spread sheet. MxFLS fieldwork protocol established to use as many supplements as needed to record all the information provided by the interviewer. The next table shows the information about the supplements:

Supplement	Sections with supplements
Book C	LS
Book 2	SU, SUA, SUG, SUF, NNA, PR, VLH
Book 3A	IE, HM, MG, MT, VLI
Book 3B	CE, HS, CR, TH, THI, TO

Book 4	HE
Book 5	EDN, CEN, HSN
Book proxy	CR, TH, THI

## 4.2 Data File naming

Data files in MxFLS follow the same structure of the questionnaire. File naming allows the user to identify each data file with its corresponding questionnaire's book and questionnaire's section. The first set of characters of the file's name identifies the questionnaire's book. The second set of characters --which follows an under-score -- identifies the section within each book.. For example the data file *iiia\_hm* corresponds to the information of book 3A and section HM of the questionnaire. Multiple data files have been created, when a particular section of the questionnaire is very large. For example, section HE (Pregnancy history) of Book 4 collects information about lifetime pregnancies. The data are very detailed and include information about health care before, during and after childbirth; data about the baby's health, and information regarding baby feeding practices. In this case, Because the information is very large, three data files are associated with this section. The number '1' in *iiiv\_he1*, denotes the file is the second data file (out of three) that contains information from book 4, section HE. The following table lists the different books that contain the the household questionnaire and their corresponding data file...

Book C	c_xx	*xx stands for the characters identifying the specific section
Book 1	l_xx	
Book 2	ii_xx	
Book 3A	liia_xx	
Book 3B	liib_xx	
Book 4	iv_xx	
Book 5	v_xx	
Book S	s_xx	
Book P	p_xx	
Book EA	ea_xx	
Book EN	en_xx	

## 4.3 Variable Naming

Variable names in the MxFLS-1 reflect the questionnaire section as well as the question number and, when necessary, the sequence number of the question. For example, the names of variables from the GH section (tastes and habits) begin with gh and end with the specific question number. For example consider question gh02 in section GH (tastes and habits) of Book 3B. This question asks for the beverages that the respondent usually drinks while eating. There are 6 options (water, soda, beer, tequila/pulque, hot drinks and other) and the respondent is allowed to give more than one answer. In this case, one variable (e.g. gh02\_1a for the first option, gh02\_1b for the second and so on) was created for each option so that the new variables contain 1 if the respondent answered affirmatively and missing value when the answer was negative.

A number of questions have two associated variables: one variable indicating whether the respondents answer the question and the "main" variable providing the respondent's answer. The fist types of variables are named by adding "\_1" to the associated question number. For example, question ed15 asked at what age did the respondent stop attending school. Variable ed10\_1 indicates whether the respondent was able to answer the question. Variable ed10\_2 provides the age.

The codebooks contain a complete list of variables by book and section together with an explanation of the codes used in every question.

#### **4.4 Identifiers and Level of Observation**

The data was organized, wherever possible, so that the level of observation within a file is either the household or the individual. When the level of observation is the household, variable FOLIO is enough to identify an observation. If the level of observation is the individual, both FOLIO and LS are required to identify a person.

In some cases the level of observation is something other than the household or individual. This is usually because the data was collected as part of a grid, in which a set of questions was repeated for a series of items or events. In these cases, data collected as part of a grid are organized rectangularly. For example, section HM (Marital history) of book 3A contains data about different number of marriages (from 1 to 8 marriages) from 14,083 individuals (not all the individuals report information about 8 marriages; this can vary from 1 to 8). Thus, there are 15,565 observations in the data file. To uniquely identify an observation in this file, the analyst should use variables FOLIO, LS and *SECUENCIA*. *SECUENCIA* is a variable that numbers the events (for a better explanation see the next section). So in this example, *SECUENCIA*=1 will stand for all first marriages reported by the individuals, while *SECUENCIA*=8 will stand for just the eighth marriage of those who reported it. Another example is found in section HE (Pregnancy history) of book 4 where every woman between 14 and 49 years provide information about their pregnancy history. In this case, the level of observation is not the individual but the pregnancies which are identified using the *SECUENCIA* variable.

#### **4.5 Special Features of the MxFLS-1 Household Data**

This section discusses the particular characteristics of MxFLS-1 household data that can affect the analysis.

##### **4.5.1 Symmetric Information**

In two MxFLS-1 sections of Book 3A, AH (assets) and DH (decision-making), husbands and wives provided symmetric information. That is, the husband answered questions about himself and about the rest of the family (including his wife); and the wife answered the same questions about herself and about the rest of the family (including her husband). These data allow comparisons of couples' perceptions of themselves and the other members of the family.

Section AH (book 3A) was only answered by married respondents whose spouse had already provided information about household assets in book 2 (whichever spouse answered book 2 did not have to answer section AH in book 3a). Within a household, if two individuals are married to each other, their AH data must be the same and the information of one person could be used to fill a gap of the other. Similarly, in section DH (book 3A) married individuals describe the way decisions are taken at the household (food, education and clothes of children, work decisions, etc). Information of both respondents could be used to check for consistencies and, in case there is lack of information of one individual, to fill the gap.

The same happens in section TH (sibling) of book 3B where detailed information about non-resident siblings is collected. We did collect information on siblings' characteristics from all respondents in the household even when the information had already been provided by another respondent. For example, if person X was a sibling of the household head, he or she had the same siblings as the household head. In that case, information was collected from both, the household head and person X.



#### 4.5.2 Duplicate Information

Some data was collected in more than one place. In most cases, the respondent was one source of information and a proxy respondent (or preprinted information) was the other. For example, book 2 about household economy (section AH) contained information on a number of topics that were also in the questionnaire books addressed to individuals. Though it would be easier to use the information from the household book, data from the individual books are likely to be more accurate, since the information was self-reported rather than provided by proxy.

**Age.** Data on age was collected in both the LS section of Book C (generally by proxy) and on the covers of the individual books. In addition, in certain places in the questionnaire, interviewers were required to examine the age recorded on the book cover, usually to determine whether the respondent was above a certain threshold age. We did not correct inconsistencies between the roster and book covers, but we created a “best-guess age” variable. We did not attempt to correct inconsistencies between the roster and questions within the book, since complicated skip patterns were often involved.

**Sex.** Information on gender was collected in the LS roster of Book C, on section SA of Book S and on the cover of books 3A and 3B. Just like with age, we did not correct inconsistencies among books. However, a “best guess sex” can be found in the in BESTGUESS.

**Marital Status.** Marital status was asked in the LS roster and on the covers of books 3A, 3B, and 4. Various interviewer checks within the individual books required using marital status information from the book cover. In cleaning the data, we tried to make sure that marital status in the roster matched marital status on the book covers. We did not clean interviewer checks because that would have required complicated adjustments to skip patterns.

**Education Level.** In the LS roster the highest level of schooling is reported and the highest grade completed within that level. For some respondents that information is repeated in book 3A or book 5. The BESTGUESS file also contains a “best-guess schooling” variable.

**Earnings and Non-labor Income.** Section TB of Book 3A asked in depth questions about employment and labor earnings. These topics are also addressed in the proxy book. TB data are likely to be more accurate because earnings were addressed in the context of related questions. Book 2, section IN, asked about non-labor income at the household level, and book 3A, section IIN, asked about it at the individual level. The individual-level IIN information is preferred, but the household summary is useful for computing total household income if an individual book is not available for all adults.

**Parents’ Survival Status.** The LS roster registered the line number for each individual's mother and father (Is06 and Is07). Special codes were used to designate whether the parent was alive but was not a member of the household; special codes were used in case a parent was dead. In book 3B, section TP (parents) explicitly asked the respondent about each parent's survival status. The TP data is considered more accurate.

#### 4.5.3 Family Relationships

The MxFLS-1 contains broad information on family relationships, mainly between husbands and wives and between parents and children. The information is not limited to household members, it also covers non-resident kin.

##### Family relationships in the LS Roster (Book C)

The LS roster provides a lot of information on relationships among current household members. The information given by the LS roster is as follows:

Variable	Information	Remarks
Is05	Which member is the household head in MxFLS-1 and how other members were related to that person.	This information indicates how members other than the household head were related. For example, if individuals 3 and 4 were both children of the head, they were either full or half-siblings. If person 5 was the mother of the head and person 3 was the child of the head, person 5 was almost certainly the grandmother of person 3. In other cases the information is not definitive. For example, if persons 6 and 7 were both grandchildren of the head, they were likely to be siblings or cousins.
Is06	LS number of an individual's father.	These variables allow linking an individual with his/her parents by giving their individual ID number (LS). For example, if person X has Is06=1 and Is07=2, his/her father is the household head (because household head always has LS =1) and his/her mother is the woman with LS=2. If a person's mother or father was death, or was alive but not a household member, especial codes were used.
Is07	LS number of an individual's mother.	
Is08	LS number of the person who takes care of children younger than 15.	

### Children Identified in Other Sections

Sections HE1 and HE2 of Book 4 contain information about the pregnancy history of all women between 14 and 49 years. The information collected includes details about the child's characteristics at the moment of birth and about the health care of the mother before, during and after the childbirth.

These sections identify each pregnancy with the identification number (Is) for each child who is a household member, allowing to join information of the child's birth with other information of the child contained in the survey. The children whose 'Is' is unknown were codified with a '99'.

**Non-resident Children.** Section TH1 in book 3B, provides the most detailed information about non-resident children. We did not collect information on children's characteristics (members of the household) if the information had already been provided by another respondent. For example, if person X (a man) was married to person Y (woman) and had had no other wives, his children were also the children of person Y. In that case, information provided by person Y about her children serves as children information for person X as well.

If person X had additional children with another woman, however, he was asked to report about those.

**Non-resident Siblings.** Section TH (siblings) in book 3B provides the most detailed information about non-resident siblings. In this case, and contrary to what we did with Non-resident Children, we did collect information on siblings' characteristics from all respondents in the household even when the information had already been provided by another respondent. For example, if person X was a sibling of the household head, he or she had the same siblings as the head. In that case, information was collected from both, the household head and person X.

## Identifying an Individual's Closest Relatives

To calculate the total number of children, siblings or living parents for a respondent, or to obtain information on the characteristics of these kin, it is necessary to merge information from several sections. The next table shows how to do so.

Sources of Information for a Respondent's Closest Relatives	
Siblings	
In the household	<ul style="list-style-type: none"><li>For a household head, use variable ls05 of section LS (book C) to identify household members who are brothers and sisters. For any other member whose mother or father is a household member, check the roster for other individuals in the household who identify the same parents in ls06/ ls07 of section LS (book C). For a non-head member whose parents are not household members, sibling information is unclear.</li><li>Once the household members who are brother or sisters have been identified, it is possible to consult their information from their individual books.</li></ul>
Outside the household	Use the respondent's TH (sibling) data from book 3B.
Children	
Of female respondents	For children ever born, use HE data from book 4. For children alive it is also possible to use book 5.
Of male respondents	For children with current wife, use information provided by wife in section HE of book 4, and book 5. For non-resident children older than 15 born to a previous wife, use THI (child) data from book 3B.
Parents	
In the household	According to the information needed, use data from books 3A and 3B completed by each parent or from section LS of book C.
Outside the household	Use the respondent's TP (parent) data from book 3B.

## 5. Community Survey

### 5.1 Survey Instruments

The community survey questionnaires were divided into books and subdivided into topical sections. The first book collected information about the general physical, economic and social characteristics of the community; the second one about prices in the community; a third one collected information about infrastructure and quality of schools; the fourth one about health service infrastructure and the last one about small health providers.

#### ***Book of Community Characteristics***

This book asked about general characteristics of the community such as population, transportation, infrastructure, credit opportunities and industries in the community. The book was answered by the president of the municipality when possible. When he or she could not be reached, the book was answered by another local authority.

**Population.** Section PB collected information about the population characteristics such as size, number of poor households, number of dwellings in the community and increase (or decrease) of the population during the last five years.

**Social aspects of the community.** Section ASC asked about the main social problems that the residents of the community may face like vandalism, people drinking or using drugs in the streets, prostitution, policemen watching and air quality.

**Natural disasters.** Section DN collected information about natural disasters that happened in the community during the last five years. Information about the exact date of the disaster and the amount of population affected was collected.

**Infrastructure history.** Section HI asked about the infrastructure changes that have occurred in the community during the last five years and how they have affected the welfare of the community. Questions were asked about construction of highways and roads, sewerage, introduction of drinking water, construction of health services and construction of new schools.

**Social Assistance.** Section AS asked about the social programs and supports that the community has received during the last 12 months. Information about the type of program and amount is collected, as well as which program has had the greater impact.

**School history.** Section SED collected information about the presence of elementary schools, secondary, high school and technological centers in the communities.

**History of health facilities.** Section SM asked about the presence of health facilities. For every health facility in the community, we collected information about the number of days a week it gives service and distance from the city center.

**Community activities.** Section AC collected information about activities in the community such as meetings or assembles. We collected information about the number of residents who participate in these activities.

**Transportation.** Section MT asked about the presence of transportation facilities and its quality (number of roads and type of materials used to cover most of them).

**Infrastructure.** Section INF collected information about the characteristics of the mass media and the infrastructure used by the inhabitants of the community such as the nearest bus/public transport

stop, the nearest foreign transport terminal, the nearest permanent market, the nearest movable market, the nearest post office or the nearest telephone. For each facility, we asked for the distance between the downtown and the facility, the main transport used to reach it, the transportation cost and the number of days a week the facility gives service. Section EL asked about the availability of electricity in the community. Section FA asked about the sources used for the community to get water for cooking and drinking and the presence of sewerage.

**Credit opportunities.** Section OC collected information about the saving habits in the community such as the institutions used to save or borrow money, its location and types of services available at these institutions.

**Industry.** Section IN asked about the industries or productive activities and its relative importance. Section EF collected detailed information about the most important productive enterprises and factories such as types of goods produced, number of workers and the average daily wage.

**Community welfare.** Section BC collected information about community welfare from different perspectives such as infrastructure, stability of prices, changes in crime level, violence, quality of education and electricity services, access to drinking water and life expectancy. Section OD includes information reported directly by the interviewer.

**Migrant clubs.** Section CM asked about assistance from migrant clubs to individuals that want to migrate or have migrated to the United States. Detailed information about the migrant clubs such as contact between the community and the clubs, size and kind of assistance provided was collected. This section also includes information about the migration patterns of the community such as migration and return rates and income transfers from migrants.

### ***Schools' Questionnaire***

This book contains detailed information about schools. It was answered by the principal of the school; in case the principal was not present, the book was answered by another authority of the school. Section PRM (questions to teachers) was answered by two teachers.

**School's principal.** Section DE collected information about the education level and activities of the school's principal such as his/her education level, the highest grade completed, how long he/she has been working in the school, whether he or she is studying *Carrera Magisterial*, the number of hours worked per week, salary and whether he/she has another job.

**School's function.** Section FE asked about the characteristics of the school educational system. The information included language used in classes, number of days worked per week, technical workshops, presence of extracurricular activities such as arts and crafts, sports, languages or computer classes, presence of tutorial or advisory teachers, types of orientation offered to students, number and quality of teachers, activities hiatus and improvement of education quality during the current year. This section also includes information about the way some changes have affected or favored the performance of the school in the previous school year such as changes in the resources of the school, in prices of didactic material (including text books), the rate of youth criminality, drug use and alcoholism in the community, domestic violence and disintegration in students homes and the political and social conflicts in the community.

**Students and teachers.** Section AM collected information about the characteristics of students and teachers of the school. The information gathered included number of groups, mobility of teachers, performance of students such as number of students registered for each grade, number of students who failed the previous year and mathematics and GPA average score by grade. The information about the teachers included highest education level completed and performance in *Carrera Magisterial*.

**Education costs and school support.** Section CA asked about the schools resources, expenses made by the students and scholarships that students from the school have access to. Information

about the expenses made by students included registration/tuition, fee/voluntary payment, average annual expenses in supplies and school books and other expenses. Finally, information about the scholarships included types of scholarships granted to students, number of students receiving scholarships and the average annual scholarship amount.

**Infrastructure.** Section IN collected information about the school facilities such as electricity, number of classrooms and laboratories, library, auditorium, long distance learning system, sport areas, quality of tables, chairs, desks, general health conditions, classroom's equipment and bathrooms.

**Questions to teachers.** Section PRM was answered by two teachers and collected information about their degrees, activities and opinions. Information about the degrees included highest educational level and grade completed and position in *Carrera Magisterial*. Information about opinions included level of domestic violence and migration in students' homes and parents' interest in their children's education.

### ***Book of Health Service Infrastructure***

This book collected detailed information about the Health Service Infrastructure. Sections RS and AG were answered by the health unit director. In case he/she was not found, whoever is in charge. For the rest of the sections, the director (or person in charge) was asked to provide the name of the persons who can give the needed information.

**Responsible for the services.** Section RS asked the director (or person in charge) to provide the name of the individuals who can give the needed information for the rest of the sections.

**General Aspect of the unit.** Section AG collected information about the general aspects of the health unit such as type of institution, identification code, whether the unit is certified, the time it has been in operation, equipment in the unit (waiting room, medical exploration area, recovery room, labor room, emergency room, operating room and blood bank), the number of cases seen last month and in the last 6 months for different services (malnutrition, cancer, diabetes, etc) and about the major social problems faced by the town such as alcoholism, drugs addiction and violence. It also collected information about the director or person in charge of the unit such as age, highest level of education, hours worked per week, wage and antiquity.

**Personnel and services provided by the unit.** Section PS asked about the personnel and facilities in the unit. Information about personnel included type and number of specialists working in the unit, while information about facilities included questions related to electricity, water sources, number of toilets, sewerage and fumigation services. Section SER collected detailed information about the services provided by the unit such as general consultations, pregnancy tests, labor care, dental exam, x ray and ultrasound. When a service was provided, information about the maximum and minimum prices was gathered. For those units with laboratory for clinical tests, section LB collected detailed information about the tests done in the laboratory and their approximate prices.

**Drugstore/medicine.** For those units with a drugstore, section FM asked about the medicines that are offered to the population. Detailed information is collected about the medicines sold, its presentations and prices.

**Direct observation.** Sections LB, RM and SE provided directly observed information about the laboratory for clinical tests, the exploration and the waiting room, respectively. The information gathered includes the interviewer's perception about the natural ventilation, air conditioning, natural and artificial light, quality of equipment and general hygiene conditions.

### ***Book of Small Health Providers***

This book collected information about the characteristics, services and materials used for the small health providers such as physician's offices, drugstores, medical clinics, midwives and other traditional practitioners. It was answered by the person in charge of providing the health service.

**General data of provider.** Section GP asked about general characteristics of the small health provider such as working hours, educational level, seniority, salary, types and number of employees and quality of services.

**Problems when offering services.** Section POS collected information about the main problems faced by the small health provider that could affect its services such as changes in prices or availability of medications, number of staff, price of water and price of electricity. Section ACG asked about social problems such as alcoholism, tabaquism, drug addiction, child abuse, domestic violence, crime and other type of injuries.

**Services offered.** Section SV asked about the services provided such as consultations to check out child nutrition, for lung tuberculosis, for high blood pressure, for diabetes, for cancer, for prenatal care, ultrasound, dental care and x-ray. For those services provided, information was collected on prices charged and frequency of visits during the last month and the last six months. Questions were asked about services of preventive medicine and family planning methods offered.

**Medicine, health material, equipment and set of instruments.** For small health providers, section MED asked about the medications that are offered to the population. Detailed information is collected about medications sold, its presentations and prices. Section MS collected information regarding the materials, equipment and set of instruments available such as general health material (antiseptics, bandages, disposable gloves, gauzes, suture thread and syringes), incubator, baby scale, dissection equipment, vaginal speculum and catheters. We also collected information about the instrument, material or equipment more frequently used and the sources used to finance them.

**Traditional midwife.** For traditional midwives, section PT collected detailed information such as training of the person in charge of providing the health service, frequency of the services and general practices.

**Medical exploration and waiting room.** Sections SRM and SA provide information about the medical exploration and the waiting room, respectively. The information gathered includes the interviewer's perception about the natural ventilation, air conditioning, natural and artificial light, quality of equipment and general hygiene conditions.

### ***Book of prices in the community***

This book collected detailed information about the prices in the community. The products reported included food, cleaning products, clothes, agricultural products and medications. Section PM (market prices) collected information of three different sources: small stands, stores, supermarkets; established and temporary markets; and others. Section PA (agricultural products) collected information from stores, office of SAGARPA, markets or other. Section PRF (prices at drugstore) collected information from one drugstore in the community. If necessary, several sources were visited to complete all prices.

**Prices in the town.** Section PM recorded detailed information about market prices of vegetables (tomato, white onion, carrots, lettuce, potato, lemon); fruits (orange, banana and red apple); different kinds of meat (chicken, pork, beef, fish); other food products (white bread, cakes, wheat flour, cooking oil, instant coffee, etc.); cleaning products (detergent, bleach, soap, toilet paper, toothpaste and broom with plastic bristles); clothing (men's shirt, woman blouse, jeans for men and women, cotton child trousers, shoes and socks); kitchen's utensils (plates, steel spoons, pewter frying pan, pewter pot, medium griddle, plastic buckets and blender); and oil fuels (gas, diesel and engine oil).

**Prices of agricultural products.** Section PA asked about prices of agricultural products such as chemical and organic fertilizer, pesticides, herbicides, tractors, yoke, seeders, pulling carts, meat bovines, dairy cows, stud bulls, pigs, cart horses, mules, sheep, goats, hens, turkeys, rabbits, fattening foods, food for hogs and sorghum.

**Prices at drugstore.** Section PRF collected information about medication prices such as antibiotics, medicine for high blood pressure, for diabetes, for tuberculosis, for malaria, diuretics, anti-parasites, analgesics, vitamin tablets and solutions. See appendix G for detailed information about the medications included in this section.

## 5.2 Data File naming

File naming convention for the community data files are the same as the ones used for the household data files. The first characters identify the associated questionnaire book, followed by characters identifying the specific section and a number denoting sequence if data from the section is spread across multiple data files.

The next table shows the name of each book and the corresponding abbreviations for the sections of that book.

Schools' Questionnaire	esc_xx	*xx stands for the characters identifying the specific section
Book of Community Characteristics	loc_xx	
Book of Small Health Providers	peq_xx	
Book of Prices in the town	pre_xx	
Book of Health Service Infrastructure	inf_xx	

## 5.3 Variable Naming

Variable naming in the community books of MxFLS-1 reflect the questionnaire section as well as the question number and, when necessary, the sequence number of the question; as aforementioned in section 4.3.

The community codebooks contain a complete list of variables by book and section together with an explanation of the codes used in every question.

## 5.4 Identifiers and Level of Observation

Generally the data has been organized so that the level of observation within a community-level file is the community. When this is the case, variable FOLIO is enough to identify an observation. Just like with the household survey, sometimes the community data was collected as part of a grid (defined above). In this case it is necessary to use both FOLIO and SECUENCIA to uniquely identify an observation.

There are other two variables that can be used as identifier: id\_loc and CLAVE. The first variable is used as a community identifier and the second one as a facility identifier. It is important to mention that variable id\_loc was created internally and does not correspond to the community identification number used by INEGI.



## 6. Linking MxFLS-1 Household Data files

MxFLS-1 data is stored in separate data files for easy handling (for example, ii\_ah of book 2 contains information about assets and ii\_in, also of book 2, about non labor income). To create analytic files, the analyst usually needs to mix data from different files. How the data should be combined depends on the nature of the desired analytic file. In this section we discuss some of the various ways to link data across files.

### 6.1 Concatenating Data

Sometimes the analyst may need to pool observations by concentrating two data files. For example, iii\_a\_ata (book 3A) and v\_atn (book 5) both contain data on time allocation. The data in iii\_a\_ata relates to adults, and the data in v\_atn relates to children. The variables for adults begin with ata, while the variables for children begin with atn, but otherwise the information is the same. It may be useful to combine the data for the two age groups, rather than keeping it in two separate files. The data can be combined into one file using the APPEND statement in STATA or the SET statement in SAS. The resulting file will contain both the observations for children and the observations for adults. Note that since variable names are different, the variables in one file should be renamed so that they match the names in the other file.

In principle, many files can be concatenated. As a general rule, when using data from books 3A, 3B and 4, check whether a corresponding section was included in the proxy book, so that the data from respondents who answered for themselves can be combined with data collected by proxy for other individuals.

The next table lists additional combinations. Keep in mind that it may be possible that some files will need to be restructured before they are linked to account for different levels of observations. Also, some files will need to have variables renamed.

**Household Files Suitable for Concatenation**

Topic	Files	Respondent Types
assets	ii_ah and iii_a_ah	Book 2 and Book 3a respondents
Credit	ii_crh and iiib_cr	Book 2 and Book 3a respondents
non-labor income	ii_in and iii_a_iin	Book 2 and Book 3a respondents
violence and crime	ii_vlh and iii_a_vli	Book 2 and Book 3a respondents
education history	iii_a_ed and v_edn	Adults and children
labor decisions	iii_a_tb and v_emn	Adults and children
time allocation	iii_a_ata and v_atn	Adults and children
outpatient utilization	iiib_ce and v_cen	Adults and children
	iiib_es and v_esn	Adults and children
health condition	v_he and v_esn	Children and children
inpatient utilization	iiib_hs and v_hsn	Adults and children

self treatment	iiib_ats and v_autn	Adults and children
Cognitive Ability	ea_eca and en_ecn	Adults and children

## 6.2 One-to-One merges at the Individual, household, facility or community level

One of the most basic linking involves linking records for a given person. If both files contain data at the same level of observation, the linkage will be a “one-to-one” merge.

**Merging Two Files at the Individual Level of Observation.** Suppose the analyst's objective is to create a file that contains information on an individual's literacy and his or her height and weight measures. The files `iiia_ed` (adults) of book 3A and `v_edn` (children) of book 5 contain information on whether respondents can read and write. The file `s_sa` of book S contains information on the respondent's height and weight measures. The three files contain one observation per individual. To create the new file sort each of the three files by folio and `ls` and then merge by folio and `ls`.

For example, suppose the analyst wants to create a file with the present weight of children and the weight at the moment of birth. The file `s_sa` of book S contains the information about the current weight, while the `iv_he2` contains the weight at the moment of birth. In both, `s_sa` and `iv_he2` there is one observation per individual; however, while in `s_sa` the observations are identified using folio and `ls`, in `iv_he2` they are identified using folio and `he06` (remember that section HE of book 4 asks about pregnancy history of a woman and identifies each pregnancy with the identification number, `ls`, for each child who is a household member). First, it is necessary to drop the adults from the `s_sa` file (remember that this file contains information about both adults and children). This can be done by dropping all those observations with an age older than an appropriate limit (this limit depends on the analyst's interests). Then, it is necessary to change the name of `he06` for `ls` in `iv_he2` (or `ls` for `he06` in `s_sa`), sort the two files by folio and `ls` and then merge by folio and `ls`. Keep in mind that since the universe of observation of `s_sa` is bigger than the universe of observations of `iv_he2`, the merge will not be perfect. There will be observations in `iv_he2` that will not be matched by any observation in `s_sa`. These observations must be dropped.

**Merging Two Files at the Household Level of Observation.** For two household-level files, such as `ii_in` and `ii_inr` of book 2, which contain data on housing characteristics, sort each file by folio and merge by folio.

**Merging Two Files at the Facility Level of Observation.** For two facility-level files, such as `esc_dg` with data on general characteristics of the school and `esc_de` with data about the principal of the school (both from the schools questionnaire); sort each file by folio and merge by folio.

**Merging Two Files at the Community Level of Observation.** Suppose the analyst wants to create a data file with information on both, general characteristics of schools contained in `esc_dg` and general characteristics of small health providers contained in `peq_gp2` for each community. In this case, previous to merging the two files, it is necessary to merge by folio each file with the cover of its corresponding book. This is important because it will add a new variable, `id_loc`, which is the community identifier and which will be the variable used in the merge. Once we make the first merge, all we have to do is sort each file by `id_loc` and merge on `id_loc`.

## 6.3 One-to-Many Merges

Frequently, it will be necessary to merge files that are not organized at the same level of observation. It is possible for a merge like this to be straightforward. However, sometimes it will require restructuring at least one of the data sets. When thinking about how to merge MxFLS-1 data

files, it is helpful to determine whether the identifying variables in one of the files are a subset of the identifying variables in the other file.

For example, suppose you want to create a file with information about education level and household assets of adult respondents. The identifying variables for the file of education (*iiia\_ed*) are *folio* and *ls*, but the identifying variable for the information about household assets (*ii\_ah*) is just *folio*. In this case, because the identifying variables in *ii\_ah* are a subset of the identifying variables in *iiia\_ed*, you could simply merge by *folio*. This will yield individual observations where the information on household assets is repeated for all the individuals residing in the same household.

Now suppose the analyst wishes to merge information on literacy (book 3A) with information on individual crime and victimization (book 3A). The identifying variables in *iiia\_ed* are *folio* and *ls*. The identifying variables in *iiia\_vli1* are *folio*, *ls* and *SECUENCIA*. In *iiia\_vli1*, an individual has as much records as assault, robberies or violent incidents reported, one for each event. The data can be merged in following ways.

First, because the identifying variables in *iiia\_vli1* are a subset of the identifying variables in *iiia\_ed*, you could simply merge by *folio* and *ls*. This yields as many records as crime incidents reported for each individual. Each record contains information about the individual's literacy and information about a particular victimization.

The second option is to restructure *iiia\_vli1* so that it is organized at the level of the individual rather than at the level of the violent incidents reported, and the identifying variables are *folio* and *ls*. This would involve creating a file that contains variables *vli(01) – vli(15)* for victimization 1, i.e., *vli12\_1\_1*, *vli12\_1\_2*, ..., *vli12\_1\_15* (in this this case is 15 because the maximum number of victimizations an individual reported is 15). This file would have more variables than *iiia\_vli1* but fewer observations. If the data from the *iiia\_vli1* file are restructured to be at the level of the individual, merging the restructured file by *folio* and *ls* with the *iiia\_ed* data yields one record per person that contains literacy information and all information on victimization.

Finally, it is worth mentioning that restructuring data files so that they are organized at a different level of observation can be done relatively easily in STATA with the *RESHAPE* commands, or in SAS with *PROC TRANSPOSE*.

## 6.4 Merging Household data with community or facility data

When merging data from the household survey with data from the community survey two main cases arise: (1) household data with community-level data and (2) household data with facility-level data.

**Merging household data with community-level data.** Suppose the analyst wants to create a data file with information on children's health status contained in section *v\_esn* of book 5, and prices of medications in the child's town contained in section *pre\_prf2* of book of prices in community. First, we have to merge *v\_esn* with the cover of book C by *folio* to add the community identifier *id\_loc* (which is only located in the cover of the C book). We have to do the same with *pre\_prf2*, that is, to add the community identifier that will allow merging the two files. Consequently, it is necessary to merge *pre\_prf2* with the cover of book of prices in community (note that variable *id\_loc* is located in the cover of the five community books and the cover of book C of the household survey) by *folio*. Once we have added *id\_loc* in both files, we have to sort both files by *id\_loc* and merge by *id\_loc*.

**Merging household data with facility-level data.** Now suppose the analyst wants to create a data file with information on children's health status contained in section *v\_esn* of book 5, and general characteristics of the hospital where he/she has been treated contained in section *inf\_ag2* of book of health service infrastructure. The key to do this merge right is to always have in mind that variable "CLAVE" serves as unique identifier of facilities. "CLAVE", which is also located in certain household data files (*c\_ah*, *c\_sp*, *iiib\_ce1*, *iiib\_hs1*, *iv\_he2*, *p\_he1*, *v\_cen1*, *v\_hsn1*, *iiia\_ed* and *v\_edna*), is the variable that will allow merging the individual level information about the children's health with the facility level information of the hospitals. First, at the community level we have to merge *inf\_ag2* with the cover of the book of health service infrastructure by *folio* to add the variable

“CLAVE”. Second, at the household/individual level we have to add the variable “CLAVE” to v\_es. To do this, we have to merge v\_esn with v\_cen1 by folio and ls. Finally, we have to sort the two newly created (the facility one and the individual one) files by “CLAVE” and merge by “CLAVE”. This will create a data file with information at the individual-level about the children’s health and the general characteristics of the hospital where he has been treated.

## **7. Cleaning the MxFLS-1 Data**

This section describes the procedures carried out during the fieldwork period and afterwards to minimize errors in the MxFLS-1 data.

### **7.1 In the Field: CAFE Editing, Interviewer Rechecks**

Data cleaning began in the field. Interviewers filled out the paper questionnaires while in the respondents’ households, then edited their work at base camp. Interviewers were responsible for handing in legible questionnaires that had been filled out as completely and accurately as possible.

A process of Computer-Assisted Field Editing (CAFE) was used to help maintain data quality in the household data. Interviewers handed in their completed paper questionnaires to a CAFE team at base camp. The CAFE team entered and edited the data on laptop computers, using data-entry software designed to detect a variety of fielding errors. Cross-book checks identified more complex inconsistencies. For example, if the gender listed for a respondent in the LS household roster was inconsistent with the respondent’s gender recorded on the cover of book 3A, an error message was generated.

The CAFE editor was responsible for discussing error messages with the interviewer. Some errors could be resolved fairly easily. For example, the interviewer might remember the gender of a respondent interviewed earlier in the day and verify that the inconsistency was due to a careless error. Other errors required the interviewer to return to the household and check with the respondent. For example, if the respondent declared in CR11 of section CR (Credit) having borrowed 7 times during the last 12 months but the detailed questions recorded only six credits, the interviewer might need to go back to the household to determine whether a seventh credit had occurred and if so, collect the information.

### **7.2 Double Data Entry and Verification**

We followed a standard procedure for eliminating transcript errors by entering the data from the paper questionnaires twice and then comparing the two sets of data. For household data, the work of the CAFE team served as the first entry; the second entry was done in INEGI offices. The two electronic versions of the data were compared and all discrepancies manually verified against the paper questionnaire. If an error occurred in the version of the dataset that was to serve as final product, the data was corrected.

In CIDE and UIA we did additional cleaning to correct remaining errors and to make the publicly available files as easy to use as possible.

### **7.3 “Look Ups”**

For detecting and resolving more complicated errors, we implemented a “Look Ups” (LU) cleaning process. LU involved the use of sophisticated, customized computer programs to run checks, with follow-up of suspected errors by specialists with extensive field experience, who

consulted the paper questionnaires. There were 3 persons working on the household survey lookups and related activities. The LU phase was important for quality assurance because the paper questionnaires sometimes contained valuable written information that was not captured in the electronic data. For example, an inconsistency might be generated because an editor made an inappropriate correction; reference to the interviewer's original annotation resolved the issue so the data could be corrected.

The LU program ran checks within and across questionnaire books for a particular household or a particular locality. As examples, the LU program checked that parents were at least 12 years older than their children.

For each error message generated, it was required to check the problem on the paper questionnaires and record it in a log file whether a correction was in fact made. If the specialist was not very sure how to correct the data, the data was not to be changed but a suggestion could be entered in the log file. Some problems were relatively straightforward to correct. Others, such as skip patterns that were not followed, could not be corrected because the data had not been collected.

The only cases corrected were ones for which positive evidence existed for a correction.

## 7.4 Special Cleaning for Open-Ended, "Other," and Numeric Variables

**Open-Ended Variables.** The questionnaire used open-ended responses for some questions that did not allow for closed-ended responses. The cleaning was done by a specially trained team at UIA and CIDE.

**Variables with "Other" Answers.** "Other" answers occurred when a response varied from the pre-coded options. In cleaning "other" responses, it was necessary to review the text responses and decide whether a response could be coded into an existing category, whether creation of new category was required, or whether the response should remain coded as "other." The cleaning was carried out by specially trained teams at CIDE and UIA.

New categories were typically created if a response was substantively different from the pre-coded responses, if it occurred in a non-trivial number of times. We were inclined to create new categories rather than leave a large "other" category. Therefore, users have the option of aggregating the data, whereas finer desegregation of the data would be impossible if new codes were not created. Subsequent to the addition of new "others" categories into the data, the questionnaires were also revised to add these new categories.

Three types of "other" variables were cleaned:

- Simple questions allowing only one answer. For this kind of questions, when pertinent, "Other" responses were recoded to a new or existing response category. For example, question 11 of section CV of book C asked what the respondent used to disinfect water. The original questionnaire (as fielded) allowed 5 substantive categories plus an "Other" option. After the cleaning process, one new category was added: "buy purified water".
- Questions where multiple responses were allowed. "Other" responses were recoded to a new or existing category, and the indicator that an "other" response had originally been selected was turned off. For example, question GH02 in section GH of book 3B asked what type of beverages the respondent prefers to drink while eating. The original questionnaire allowed 4 substantive categories (water, beer, soda and tequila/pulque) plus an "Other" option. After the cleaning process, two new categories were added (rum and hot drinks) and the "Other" option was divided in "Other alcoholic drinks" and "Other non-alcoholic drinks".

- Questions related to items in a grid. Cleaning of “Other” responses here might generate another item in the grid. If a new category was created, the “other” code was deleted. For example, question 39 of section HE of Book 4 asked who provided care during labor. In this case, the original questionnaire allowed 8 substantive categories (general doctor, gynecologist, pediatrician, midwife, auxiliary or health promoter, nurse or nobody) plus an “Other” option. After the cleaning process, one new category was added (anesthesiologist).

At the end of this cleaning process, the observations that remained in the “Other” categories correspond to text responses. Note that the “Other variables” is not of public use.

**Numeric Variables.** Some numeric responses did not fit the space provided, either because the answer had too many digits or required more decimal places than were allowed. In these cases, interviewers had been trained to fill the space provided with a code of 9’s ending in a 5 (“out of range”) and to record the correct answer in the “Notes” section of the questionnaire or in the “other” answers file. If justified by the interviewer’s annotations, we widened the numeric field to allow the correct answer and replaced the “out of range” code with the correct answer. It was not possible to correct all out of range codes, so *special codes* sometimes still appear in the data.

## 7.5 Section Checks

For each data section, we made an effort to:

- Review the LU checks and determine whether any remaining errors or inconsistencies could be corrected.
- Review numeric responses for the existence of special codes and review character variables for responses meaning “empty” or “don’t know”.
- Create or correct variables “\_1”, “\_2” so that the special codes were preserved and the associated numeric or character variable contained only valid responses. “\_1” variables are associated typically with a numeric value and indicate whether or not the person was able to answer the question.
- Check that skip patterns were properly followed and apply corrections if data would not be lost as a result.
- Assign variable names and labels as clearly as possible.
- Check for duplicate observations.
- Find and drop any variables that might enable identification of a respondent.

## 7.6 Checks on IDs across Books

It is essential that Identification numbers (ID) such as FOLIO and LS be correctly assigned. Therefore, we rigorously checked ID assignment by using a program that followed the same IDs through all survey sections. This program verified the age and name of each individual surveyed to ensure that there were no mistakes on ID assignment through the different sections.

## 7.7 Checks on Units of Measure

Some questions asked for a numeric answer and allowed a choice of units of measure. For example, CS10 asked the amount of several goods bought in the last 7 days, allowing for an answer either in kilograms, pieces, or other unit. Occasionally respondents provided answers that were clearly outliers. For example, if a respondent said that he/she bought 10 kilograms of bakery or store unpacked bolillo/telera (bread), we judged the proper unit to be pieces rather than kilograms. Similarly, if a woman reported a miscarriage after a pregnancy of 11 months, we judged the proper unit to be weeks rather than months. Such corrections typically involved very few cases.

## 7.8 Created Variables and Files

Some variables and data files were created to make the data easier to use. For example:

- Occupation and sector had pre-coded answers in section TB, but we also obtained open-ended answers. The open-ended answers were later coded by INEGI into CMO codes for occupation and NACIS (SCIAN in Spanish) codes for activity classifications. These new variables are named ‘...cmo’ and ‘...scian’ and are followed by the number of the question they refer to. This is extensively explained in Appendix C.
- Since age and date of birth information, as well as schooling information, can vary across books, we constructed our “best guess” of each person’s age and schooling variables using all of the data in MxFLS-1 and reported this in BESTGUESS database. This is available upon request.
- For Book 4 (Reproductive Health) we substituted children’s names that were provided by their mothers for the corresponding ID number of each child. This could only be done for children who lived with their mothers and whose mothers are 14 to 49 years old. This is the case of variable he06 contained in section HE1, and of variable he19a contained in section HE2. These variables assign a code of ‘99’ for those children we did not find in the household.

## 8. References

INEGI (2004). "Sample Design". Description of MxFLS Baseline Sample. CIDE & UIA Working Paper.

*Indonesian Family Life Survey (IFLS)*

Raven, J.C., Court, J.H., & Raven, J. (1986). Manual for Raven's Progressive Matrices and Vocabulary Scales (Section 2) - Coloured Progressive Matrices (1986 edition with U.S. norms). London: Lewis.

Raven, J.C., Court, J.H., & Raven, J. (1983). Manual for Raven's Progressive Matrices and Vocabulary Scales (Section 3) - Standard Progressive Matrices (1983 edition). London: Lewis.

NAICS (North American Industrial Classification System)

CMO (Mexican Classification of Occupation). INEGI, [www.inegi.gob.mx](http://www.inegi.gob.mx)



## Appendix A

This appendix presents the response rates of the survey and the notes on response burden.

### Response Rates

MxFLS-1 household response rates are presented on the following table:

	Rates (%)
Book I (household expenditures)	95
Book II (household economy)	95
Books IIIA and B (adults/individual)	91 (8% with proxy)
Book IV (Reproductive health, contraception and birth outcomes)	91 (3% with proxy)
Book V (Children-individual)	95
Cognitive children status	90
Cognitive adults status	85
Anthropometrics measures	91
Hemoglobin	78

### Notes on Response Burden

The household survey instrument is complicated and takes time to complete. In MxFLS-1 we attempted to organize and format the instrument to minimize response burden. The mean time to complete a book varied across the books, with the longest time observed for the household economy book and the individual-level books addressed to adults, varying depending on the number of adults the household had. The next table shows mean completion time of respondents for the different books in the survey, as well as the mean completion time for the entire survey. For the individual level books, the mean time was obtained by the mean of the sum of the completion time of each individual belonging to the household. For the household level books (C, 1 and 2) the time ranges between 22 and 25 minutes, while for the individual level books (3A, 3B, 4, 5) the time variation is larger.

Mean Time per Household by Book		
	Mean time (in minutes) per household in answering:	Mean time per household in answering the complete survey
Book C	25	5 hrs 45 min
Book I	23	
Book 2	26	
Book 3A	88.6	
Book 3B	85	
Book 4	38	
Book 5	62	
Book S	34	
Book P	42	
Book EA	44	
Book EN	26	

## Appendix B

### Names of Data Files for the Household Survey

Book	File Name	Sections Content	Level of Observation	No. Records
Book C	c_conpor	Back page	Household	50,568
	c_cv	CV- Dwelling characteristics	Household	8,440
	c_cvo	CVO- Dwelling characteristics (direct observation)	Household	8,435
	c_eh	EH- Household Members schooling	Household	6,506
	c_ls	LS- Household Roster	Individual	35,677
	c_ne	NE- Interview notes	Household	8,434
	c_portad	Cover	Household	8,441
	c_rc	RC- Information for recontact	Household	8,438
	c_sp	SP- Family planning and health	Household	14,646
Book I	i_conpor	Back page	Household	48,233
	i_cs	CS- Consumption (1)	Household	8,051
	i_cs1	CS- Consumption (2)	Household	8,050
	i_ne	NE- Interview notes	Household	8,046
	i_portad	Cover	Household	8,052
Book II	ii_ah	AH- Household Assets (1)	Household	8,046
	ii_ah1	AH- Household Assets (2)	Household	7,992
	ii_conpor	Back page	Household	48,258
	ii_crh	CRH- Household Credit	Household	8,049
	ii_in	IN- Household Non labor income	Household	8,048
	ii_inr	INR- Rural income	Household	8,046
	ii_ne	NE- Interview notes	Household	8,050
	ii_nna	NNA- Non agricultural business (1)	Household	8,051
	ii_nna1	NNA- Non agricultural business (2)	Household	1,283
	ii_portad	Cover	Household	8,051
	ii_se	SE- Household Economic shocks	Household	8,049
	ii_su	SU- Land (1)	Household	8,050
	ii_su1	SU- Land (2)	Household	2,246
	ii_vlh	VLH- Household Crime and victimization (1)	Household	8,049
	ii_vlh1	VLH- Household Crime and victimization (2)	Household	1,046
Book III A	iiia_ah	AH- Household assets (1)	Individual	19,746
	iiia_ah1	AH- Household assets (2)	Individual	5,781
	iiia_ata	ATA- Adult time allocation	Individual	19,754
	iiia_conpor	Back page	Individual	118,404
	iiia_dh	DH- Home decisions	Individual	138,288
	iiia_ed	ED- Education	Individual	19,757
	iiia_hm	HM- Marital history (1)	Individual	19,748
	iiia_hm1	HM- Marital history (2)	Individual	15,565

	iii_a_ie	IE- Schooling interruptions (1)	Individual	5,841
	iii_a_ie1	IE- Schooling interruptions (2)	Individual	227
	iii_a_iin	IIN- Individual nonlabor income	Individual	11,843
	iii_a_mg	MG- Permanent migration (1)	Individual	19,756
	iii_a_mg1	MG- Permanent migration (2)	Individual	9,196
	iii_a_mt	MT- Temporary migration (1)	Individual	19,738
	iii_a_mt1	MT- Temporary migration (2)	Individual	1,132
	iii_a_ne	NE- Interview notes	Individual	19,754
	iii_a_portad	Cover	Individual	19,764
	iii_a_shi	SHI- Individual shocks	Individual	19,755
	iii_a_tb	TB- Employment	Individual	19,755
	iii_a_vli	VLI- Crime and victimization (1)	Individual	19,755
	iii_a_vli1	VLI- Crime and victimization (2)	Individual	3,216
Book III B	iii_b_ats	ATS- Self-treatment	Individual	19,803
	iii_b_ca	CA- Insurance condition	Individual	19,800
	iii_b_ce	CE-Outpatient Utilization (1)	Individual	19,803
	iii_b_ce1	CE-Outpatient Utilization (2)	Individual	4,333
	iii_b_conpor	Back page	Individual	120,785
	iii_b_cr	CR-Credit (1)	Individual	19,802
	iii_b_cr1	CR-Credit (2)	Individual	2,687
	iii_b_ec	EC- Acute Morbidity	Individual	17,728
	iii_b_es	ES- Health condition (1)	Individual	19,804
	iii_b_es1	ES- Health condition (2)	Individual	5,104
	iii_b_gh	GH- Tastes and habits	Individual	19,804
	iii_b_hs	HS-Inpatient Utilization (1)	Individual	19,799
	iii_b_hs1	HS-Inpatient Utilization (2)	Individual	1,200
	iii_b_ne	NE-Interview notes	Individual	19,799
	iii_b_portad	Cover	Individual	19,809
	iii_b_re	RE-Recontact information of relatives in USA (1)	Individual	19,797
	iii_b_re1	RE-Recontact information of relatives in USA (2)	Individual	11,386
	iii_b_sm	SM- Emotional Wellbeing	Individual	19,803
	iii_b_th	TH-Transfers of non-resident siblings (1)	Individual	19,799
	iii_b_th1	TH-Transfers of non-resident siblings (2)	Individual	17,829
	iii_b_th2	TH-Transfers of non-resident siblings (3)	Individual	75,938
	iii_b_thi	THI- Transfers of non-resident children (1)	Individual	19,801
	iii_b_thi1	THI- Transfers of non-resident children (2)	Individual	2,714
	iii_b_thi2	THI- Transfers of non-resident children (3)	Individual	13,725
	iii_b_to	TO-Transfers of other non-resident persons	Individual	19,801
	iii_b_tp	TP-Transfer of non-resident parents	Individual	19,801
Book IV	iv_ac	AC- Contraception	Individual	8,740
	iv_conpor	Back page	Individual	52,372
	iv_he	HE- Pregnancy history (1)	Individual	8,740
	iv_he1	HE- Pregnancy history (2)	Individual	18,765
	iv_he2	HE- Pregnancy history (3)	Individual	10,463
	iv_ne	NE- Interview notes	Individual	8,739
	iv_portad	Cover	Individual	8,743
	iv_res	RES- Pregnancy summary	Individual	8,739

Book V	v_atn	ATN- Child time allocation	Individual	11,314
	v_autn	AUTN- Child self treatment	Individual	11,311
	v_cen	CEN- Child outpatient utilization (1)	Individual	11,311
	v_cen1	CEN- Child outpatient utilization (2)	Individual	2,119
	v_conpor	Back page	Individual	67,720
	v_edna	EDN- Child's education (1)	Individual	11,315
	v_ednb	EDN- Child's education (2)	Individual	6,640
	v_edn1	EDN- Child's education (3)	Individual	176
	v_emn	EMN- Child employment	Individual	11,314
	v_esn	ESN- Child health condition	Individual	11,311
	v_hsn	HSN- Child inpatient utilization (1)	Individual	11,310
	v_hsn1	HSN- Child inpatient utilization (2)	Individual	325
	v_ne	NE- Interview notes	Individual	11,309
	v_portad	Cover	Individual	11,315
	v_vac	VAC- Child vaccination	Individual	11,313
Proxy Book	p_ac	AC- Contraception	Individual	293
	p_ca	CA-Insurance condition	Individual	1,848
	p_ce	CE-Outpatient Utilization	Individual	1,848
	p_conpor	Back page	Individual	11,362
	p_cr	CR- Credit (1)	Individual	1,848
	p_cr1	CR- Credit (2)	Individual	117
	p_ed	ED-Education	Individual	1,879
	p_es	ES- Health condition (1)	Individual	1,848
	p_es1	ES- Health condition (2)	Individual	511
	p_gh	GH- Tastes and habits	Individual	1,848
	p_he	HE- Pregnancy history (1)	Individual	293
	p_he1	HE- Pregnancy history (2)	Individual	181
	p_hm	HM- Marital history	Individual	1,879
	p_hs	HS- Inpatient Utilization	Individual	1,848
	p_mg	MG- Permanent migration	Individual	1,879
	p_ne	NE- Interview notes	Individual	1,902
	p_portad	Cover	Individual	1,903
	p_res	RES- Pregnancy summary	Individual	549
	p_tb	TB-Employment	Individual	1,879
	p_th	TH- Transfers of non-resident siblings (1)	Individual	1,848
	p_th1	TH- Transfers of non-resident siblings (2)	Individual	1,422
	p_th2	TH- Transfers of non-resident siblings (3)	Individual	5,876
	p_thi	THI- Transfers of non-resident children (1)	Individual	1,848
	p_thi1	THI- Transfers of non-resident children (2)	Individual	234
	p_thi2	THI- Transfers of non-resident children (3)	Individual	1,060
	p_to	TO- Transfers of other non-resident persons	Individual	1,848
	p_tp	TP- Transfers of non-resident parents	Individual	1,845
Book S	s_conpor	Back page	household	37,905
	s_ne	NE- Interview notes	household	3,699
	s_portad	Cover	household	7,587
	s_sa	SA- Anthropometrics and biomarkers	Individual	32,169

## Appendix C:

### List of public and non public open question

#### Household data

Non public questions			
Variable	Question	Section	Book
Is01	Household member's full name	Household roster	C
Is17	What is the name of the school that (..) currently attends?	Household roster	C
rc01	If you or any member of you family moves, who would be able to give us information regarding your whereabouts?	Recontact	C
rc02	Could you give the name of any other relative or friend who would know about you in case you moved?	Recontact	C
sp01	What is the address of the hospital/clinic?	Family planning and health	C
sp02	What is the address of the small health provider?	Family planning and health	C
eh02	Name of the school?	Schooling of household members	C
eh04	What is the address of the school?	Schooling of household members	C
Ne04	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes	C
Ne05	What questions did interviewer find difficult, embarrassing or confusing?	Interview sessions notes	C
Ne06	What questions did respondent seem interested in?	Interview sessions notes	C
Ne07	Notes	Interview sessions notes	C
Ne04	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes	1
Ne05	What questions did interviewer find difficult, embarrassing or confusing?	Interview sessions notes	1
Ne06	What questions did respondent seem interested in?	Interview sessions notes	1
Ne07	Notes	Interview sessions notes	1
su05	Reference of plot/land	Land	2
nna04	Reference of business	Non-agricultural business	2
nna06	What did the business mainly do when it started?	Non-agricultural business	2
vlh20	Reference of incident	Household crime and victimization	2
Ne04	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes	2
Ne05	What questions did interviewer find difficult, embarrassing or confusing?	Interview sessions notes	2
Ne06	What questions did respondent seem interested in?	Interview sessions notes	2
Ne07	Notes	Interview sessions notes	2
ed27	What is the name and the address of school that you attend/ attended, and where is it	Education	3A

	located?		
mg01	¿ What is the name of the locality/community /municipality/district, state, and country where you were born?	Permanent migration	3A
mg04	What is the name of the locality, community, municipality/district, state, and country where you lived when you were 12 years old?	Permanent migration	3A
mg17	What is the name of the locality/community/ municipality/ district, state, and country where you arrived when you moved to [...]?	Permanent migration	3A
mg35	Where do you think you could move to?	Permanent migration	3A
mt05	What is the name of the locality/community /municipality/ state, and country, where you went on your trip to (...) ?	Temporary migration	3A
mt06	What is the name of the locality/community /municipality/state, and country, where you lived before your trip to (...) ?	Temporary migration	3A
mt04	Reference of trip	Temporary migration	3A
vli08	Reference of incident	Temporary migration	3A
Ne04	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes	3A
Ne05	What questions did interviewer find difficult, embarrassing or confusing?	Interview sessions notes	3A
Ne06	What questions did respondent seem interested in?	Interview sessions notes	3A
Ne07	Notes	Interview sessions notes	3A
ce09	Which was the reason why you went to [...]?	Outpatient Utilization	3B
ce11	What is the address of [...] where you went when [...]?	Outpatient Utilization	3B
hs09	Which is the address of [...] where you attended when [...]?	Inpatient Utilization	3B
hs07	¿ Which was the reason why you went to [...]?	Inpatient Utilization	3B
tp10	Do you know where your [...] was borne?	Non-resident parents transfers	3B
tp18	Where does your [...] /did your [...] live before dying?	Non-coresident parents transfers	3B
Ne04	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes	3B
Ne05	What questions did interviewer find difficult, embarrassing or confusing?	Interview sessions notes	3B
Ne06	What questions did respondent seem interested in?	Interview sessions notes	3B
Ne07	Notes	Interview sessions notes	3B
he24	Can you give me the name and the address of the place that you visited for check ups?	Pregnancy history	4
he32	Can you give me the name and the address of the place that you visited for childbirth?	Pregnancy history	4
Ne04	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes	4
Ne05	What questions did interviewer find difficult, embarrassing or confusing?	Interview sessions notes	4
Ne06	What questions did respondent seem interested in?	Interview sessions notes	4
Ne07	Notes	Interview sessions notes	4
edn13	What is the name and address of the school?	Child's education	5
edn33	What is the name and address of the school	Child's education	5

cen10	in 2001? What is the address where (NAME OF THE BOY/GIRL) went when [...]?	Child outpatient utilization	5
hsn08	What is the name and address of the fertility [...]?	Child inpatient utilization	5
Ne04	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes	5
Ne05	What questions did interviewer find difficult, embarrassing or confusing?	Interview sessions notes	5
Ne06	What questions did respondent seem interested in?	Interview sessions notes	5
Ne07	Notes	Interview sessions notes	5
mg01	¿what is the name of the locality/community /municipality/district, state, and country where you were born?	Permanent migration	Proxy
mg04	What is the name of the locality, community, municipality/district, state, and country where you lived when you were 12 years old?	Permanent migration	Proxy
mg09	What is the name of the locality/ community, municipality/district, state, and country where (name) lived before moving?	Permanent migration	Proxy
tp10	Do you know where (Name's) [...] was born?	Non-resident parents transfers	Proxy
tp16	Do you know if (Name's) [...] suffers/suffered any chronicle or physical illness (deafness, paralysis, blindness, etc.)?	Non-resident parents transfers	Proxy
tp18	Where does/did (Name's) [...] live before dying?	Non-resident parents transfers	Proxy
he13	Can (NAME) give me the name and the address of the place that (NAME) visited for check ups?	Pregnancy history	Proxy
he21	Can (NAME) give me the name and the address of the place that (NAME) visited for the childbirth?	Pregnancy history	Proxy
Ne04	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes	Proxy
Ne05	What questions did interviewer find difficult, embarrassing or confusing?	Interview sessions notes	Proxy
Ne06	What questions did respondent seem interested in?	Interview sessions notes	Proxy
Ne07	Notes	Interview sessions notes	Proxy

Note: questions regarding the address (neighborhood, street and number) of the household present in the cover of all books are not public.

Public questions			
variable	Question	Section	Book
su11.	During the last 12 months, which have been the three most important products / crops grown on the plot/land (...) ?	Land	2
su07, su17	Not present in the database	Land	2
nna06.	What did the business <i>mainly</i> do [...] when it started?	Non-agricultural business	2
mg01	What is the name of the locality/community	Permanent migration	3A

	/municipality/district, state, and country where you were born?		
mg04	What is the name of the locality, community, municipality/district, state, and country where you lived when you were 12 years old?	Permanent migration	3A
mg17	What is the name of the locality/community/ municipality/ district, state, and country where you arrived when you moved to [...]?	Permanent migration	3A
mg35	Where do you think you could move to?	Permanent migration	3A
mt05	What is the name of the locality/community/ municipality/ state, and country, where you went on your trip to (...) ?	Temporary migration	3A
mt06	What is the name of the locality/community/ municipality/ state, and country, where you lived before your trip to (...) ?	Temporary migration	3A
es10	Which are the three most serious health problems, you have had during your life?	Health condition	3B
tp10	Do you know where your [...] was borne?	Non-resident parents transfers	3B
tp16	Do you know if your [...] suffers/suffered any chronicle or physical illness (deafness, paralysis, blindness, etc.)?	Non-resident parents transfers	3B
tp18	Where does your [...] /did your [...] live before dying?	Non-resident parents transfers	3B
cr17	Why weren't you given the loan for [...]?	Credit	3B
cr15	Were you asked for any guarantee when you borrowed [...]?	Credit	3B
cr13	What was the reason you borrowed?	Credit	3B
edn21	For what reason did (NAME OF THE BOY /GIRL) stop attending school?	Child's education	5
edn37a	Reference of school grade	Child's education	5
emn08	Which activity did (NAME OF THE BOY /GIRL) do to help household expenditures?	Child employment	5
emn12	Which activity does/did (NAME OF THE BOY /GIRL) do to help household expenditures?	Child employment	5
sa07	Reason for not being measured	Anthropometrics and biomarkers	S
sa09	Reason for not being weighted	Anthropometrics and biomarkers	S
sa11	Why encephalic perimeter was not measure?	Anthropometrics and biomarkers	S
sa14	Why waist was not measure?	Anthropometrics and biomarkers	S
sa15	Why hip was not measure?	Anthropometrics and biomarkers	S
sa16	Why blood pressure was not measure?	Anthropometrics and biomarkers	S
sa17	Why hemoglobin was not measure?	Anthropometrics and biomarkers	S
sa18	Write down if the respondent suffers any visible illness	Anthropometrics and biomarkers	S
sa19	Write down if the respondent suffers any handicap	Anthropometrics and biomarkers	S
sa20	Specify why the measures could not be taken	Anthropometrics and biomarkers	S
mg01	¿ What is the name of the locality /community/ municipality/district, state, and country where you were born?	Permanent migration	Proxy
mg04	What is the name of the locality, community, municipality/district, state, and country where you lived when you were 12 years old?	Permanent migration	Proxy
mg09	What is the name of the locality/ community, municipality/district, state, and country where (name) lived before moving?	Permanent migration	Proxy
cr13	What was the reason you borrowed?	Credit	Proxy
cr15	Were you asked for any guarantee when you	Credit	Proxy



	borrowed [...]?		
cr17	Why wasn't (NAME) given the loan for [...]?	Credit	Proxy
tp10	Do you know where (Name's) [...] was born?	Non-resident parents transfers	Proxy
tp16	Do you know if (Name's) [...] suffers/suffered any chronicle or physical illness (deafness, paralysis, blindness, etc.)?	Non-resident parents transfers	Proxy
tp18	Where does/did (Name's) [...] live before dying?	Non-resident parents transfers	Proxy
es10	Which are the three most serious health problems, (NAME) has had during his/her life?	Health condition	Proxy
es08	Did (NAME) have a permanent injury that changed his/her way of living due to the accident?	Health condition	Proxy

## Community data

Non public questions			
Variable	Question	Section	Book
Pb02	Name, occupation and place of the respondent	Population	Community Characteristics
Pb05	What is the new name and what was the old name of the community?	Population	Community Characteristics
Pb06	What was the main reason for changing the name of this community the last time?	Population	Community Characteristics
Ef03	Could you give me the name of the three most important enterprises/factories in this community?	Enterprises and factories	Community Characteristics
Ef15	Could you give the name of the three enterprises /factories that were closed?	Enterprises and factories	Community Characteristics
Cm03	What is the migrant association/club name and address in [...] to where community members regularly attend?	Migrants clubs	Community Characteristics
Cm23	With which institution the migrants association/club located in [...] has agreements?	Migrants clubs	Community Characteristics
Ne04	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes	Community Characteristics
Ne05	What questions did interviewer find difficult, embarrassing or confusing?	Interview sessions notes	Community Characteristics
Ne06	What questions did respondent seem interested in?	Interview sessions notes	Community Characteristics
Ne07	Notes	Interview sessions notes	Community Characteristics
Ne104	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes (1)	Prices in community
Ne105	What questions did interviewer find difficult, embarrassing or confusing?	Interview sessions notes (1)	Prices in community
Ne106	What questions did respondent seem interested in?	Interview sessions notes (1)	Prices in community
Ne107	Notes	Interview sessions notes (1)	Prices in community
Ne204	What questions did respondent find difficult, embarrassing or confusing?	Interview sessions notes (2)	Prices in community
Ne205	What questions did interviewer find difficult,	Interview sessions	Prices in

Ne206	embarrassing or confusing? What questions did respondent seem interested in?	notes (2) Interview notes (2)	sessions	community Prices in
Ne207	Notes	Interview notes (2)	sessions	community Prices in
Ne304	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (2)	sessions	community Prices in
Ne305	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (3)	sessions	community Prices in
Ne306	What questions did respondent seem interested in?	Interview notes (3)	sessions	community Prices in
Ne307	Notes	Interview notes (3)	sessions	community Prices in
Ne104	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (1)	sessions	Schools' Questionnaire
Ne105	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (1)	sessions	Schools' Questionnaire
Ne106	What questions did respondent seem interested in?	Interview notes (1)	sessions	Schools' Questionnaire
Ne107	Notes	Interview notes (1)	sessions	Schools' Questionnaire
Ne204	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (2)	sessions	Schools' Questionnaire
Ne205	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (2)	sessions	Schools' Questionnaire
Ne206	What questions did respondent seem interested in?	Interview notes (2)	sessions	Schools' Questionnaire
Ne207	Notes	Interview notes (2)	sessions	Schools' Questionnaire
Ne304	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (3)	sessions	Schools' Questionnaire
Ne305	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (3)	sessions	Schools' Questionnaire
Ne306	What questions did respondent seem interested in?	Interview notes (3)	sessions	Schools' Questionnaire
Ne307	Notes	Interview notes (3)	sessions	Schools' Questionnaire
Ne404	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (4)	sessions	Schools' Questionnaire
Ne405	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (4)	sessions	Schools' Questionnaire
Ne406	What questions did respondent seem interested in?	Interview notes (4)	sessions	Schools' Questionnaire
Ne407	Notes	Interview notes (4)	sessions	Schools' Questionnaire
Ne504	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (5)	sessions	Schools' Questionnaire
Ne505	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (5)	sessions	Schools' Questionnaire
Ne506	What questions did respondent seem interested in?	Interview notes (5)	sessions	Schools' Questionnaire
Ne507	Notes	Interview notes (5)	sessions	Schools' Questionnaire
Ne604	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (6)	sessions	Schools' Questionnaire
Ne605	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (6)	sessions	Schools' Questionnaire
Ne606	What questions did respondent seem interested in?	Interview	sessions	Schools'

Ne607	Notes	notes (6) Interview notes (6)	sessions	Questionnaire Schools'
Ne704	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (7)	sessions	Questionnaire Schools'
Ne705	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (7)	sessions	Questionnaire Schools'
Ne706	What questions did respondent seem interested in?	Interview notes (7)	sessions	Questionnaire Schools'
Ne707	Notes	Interview notes (7)	sessions	Questionnaire Schools'
Rs03	What is the name of the person who can tell us about the (...)?	Responsible for the services		Health Service Infrastructure
Ne104	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (1)	sessions	Health Service Infrastructure
Ne105	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (1)	sessions	Health Service Infrastructure
Ne106	What questions did respondent seem interested in?	Interview notes (1)	sessions	Health Service Infrastructure
Ne107	Notes	Interview notes (1)	sessions	Health Service Infrastructure
Ps01	Interviewer: verify rs03 and write down the name of the person that will answer this section. If not available he / she could be replaced.	Personnel and services that the unit has		Health Service Infrastructure
Ne204	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (2)	sessions	Health Service Infrastructure
Ne205	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (2)	sessions	Health Service Infrastructure
Ne206	What questions did respondent seem interested in?	Interview notes (2)	sessions	Health Service Infrastructure
Ne207	Notes	Interview notes (2)	sessions	Health Service Infrastructure
Ser01	Interviewer: verify rs03 and write down the name of the person that will answer this section. If not available he / she could be replaced.	Activities and services provided by the unit		Health Service Infrastructure
Ne304	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (3)	sessions	Health Service Infrastructure
Ne305	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (3)	sessions	Health Service Infrastructure
Ne306	What questions did respondent seem interested in?	Interview notes (3)	sessions	Health Service Infrastructure
Ne307	Notes	Interview notes (3)	sessions	Health Service Infrastructure
Lab02	Interviewer: verify rs03 and write down the name of the person that will answer this section. If not available he / she could be replaced	Laboratory for clinical tests		Health Service Infrastructure
Ne404	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (4)	sessions	Health Service Infrastructure
Ne405	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (4)	sessions	Health Service Infrastructure
Ne406	What questions did respondent seem interested in?	Interview notes (4)	sessions	Health Service Infrastructure
Ne407	Notes	Interview notes (4)	sessions	Health Service Infrastructure
Fm02	Interviewer: verify rs03 and write down the name of the person that will answer this section. If not available he / she could be replaced	Drugstore/Medicines		Health Service Infrastructure

Ne504	What questions did respondent find difficult, embarrassing or confusing?	Interview notes (5)	sessions	Health Service Infrastructure
Ne505	What questions did interviewer find difficult, embarrassing or confusing?	Interview notes (5)	sessions	Health Service Infrastructure
Ne506	What questions did respondent seem interested in?	Interview notes (5)	sessions	Health Service Infrastructure
Ne507	Notes	Interview notes (5)	sessions	Health Service Infrastructure
Gp01	Name of respondent	General provider	data of	Small Health Providers

Note: questions regarding the address (neighborhood, street and number) of the facilities collected in the cover of all books except in characteristics of the community and prices are not public.

Public questions				
Variable	Question	Section		Book
As04	Which is the main activity of the three most important cooperatives of the community?	Community aspects	social	Community Characteristics
As14	Could you give me the name of the three programs that have the greater number of beneficiaries or the greater impact in the last 12 months [DATE OF A YEAR AGO] to [DATE OF THE INTERVIEW] in the community?	Community aspects	social	Community Characteristics
Ac04	Which are the five most important or greatest community activities that have been done in the last 12 months?	Communitarian activities		Community Characteristics
I02	By order of importance, which are the three most important economic activities that generate the income in this community?	Industry		Community Characteristics
Ef04	What are the types of products produced /services granted by [ENTERPRISE NAME/FACTORY]?	Enterprises factories	and	Community Characteristics
Ef16	What were the types of products produced /services granted by [ENTERPRISE NAME/FACTORY]?	Enterprises factories	and	Community Characteristics
Cm21	What kind of agreement exists between the state /municipality government and the migrants association/club located in [...]?	Migrants clubs		Community Characteristics
Cm24	Which kind of agreements exists between the institution and the migrants association/club located in [...]?	Migrants clubs		Community Characteristics
Fe48	What measures did the school take to respond to these events?	School's function		Schools' Questionnaire
Ca17	How were these resources obtained?	Education costs and school aids		Schools' Questionnaire
Ca19	How have the resources for the current school year been obtained?	Education costs and school aids		Schools' Questionnaire
Prm29	What type of course/workshop, different from your Carrera Magisterial, was the one that you took the last time?	Questions teachers	to	Schools' Questionnaire
Prm38	What module or course you consider should be taught to the students of this school that are not taught at the	Questions teachers	to	Schools' Questionnaire

Prm99	present time? What type of course/workshop, different from your Carrera Magisterial, was the one that you took the last time?	Questions teachers	to Schools' Questionnaire
Prm108	What module or course you consider should be taught to the students of this school that are not taught at the present time?	Questions teachers	to Schools' Questionnaire
Rs04	What position does he have?	Responsible for the services	Health Service Infrastructure
Ag55	Where are the patients referred to when they need some laboratory test?	General aspects of the unit	Health Service Infrastructure
Ag62	In your opinion, what are the three main problems that are facing this health unit at the moment?	General aspects of the unit	Health Service Infrastructure
Ag63	In your opinion, what are the 3 main problems that this unit has been facing or still is facing the last 5 years?	General aspects of the unit	Health Service Infrastructure
Ag72	What measures has this unit taken to solve this change?	General aspects of the unit	Health Service Infrastructure
Ser46	If a patient must be referred to another facility what are the two main places you send the patients?	Activities and services provided by the unit	Health Service Infrastructure
Lab06	For lab work not done here, is the patient referred outside?	Laboratory for clinical tests	Health Service Infrastructure
Fm06	What is the most sold or given away presentation of this medicine? (quantity in tablets, milliliters, grams, etc)	Drugstore/Medicines	Health Service Infrastructure
Rm08	From what you observed, write down what caught your attention.	Direct observation of exploration room	Health Service Infrastructure
Se15	From what you observed, write down what called your attention	Direct observation of waiting room	Health Service Infrastructure
Pos04	How has the change affected you?	Problems when offering services	Small Health Providers
Sv17	Specify other vaccines offered	Services offered	Small Health Providers
Sv33	Can you tell me the 5 services that you most often offer?	Services offered	Small Health Providers
Sv40	If a patient must be referred to another facility, what are the two main places you send the patients?	Services offered	Small Health Providers
Ms05	Sets of instruments, material or equipment that you generally use	Health material, equipment and set of instruments	Small Health Providers
Med06	What is the most sold or given away presentation of this medicine? (quantity in tablets, milliliters, grams, etc)	Medicines	Small Health Providers
Med11	Which are these medicines?	Medicines	Small Health Providers
Med20	Have you had any problems to get (...) in the last 12 months?	Medicines	Small Health Providers
Med21	Where do you get (...)?	Medicines	Small Health Providers
Srm12	From what you observed write down what caught your attention.	Medical exploration room	Small Health Providers
Sa16	From what you observed, write down what caught Your attention.	Waiting room	Small Health Providers

## Appendix D

This appendix presents a brief description of the NAICS (North American Industrial Classification System) and the CMO (Mexican Classification of Occupation).

### NAICS

The North American Industry Classification System (NAICS) is the new standard code system to describe business establishments and industries, replacing the Standard Industrial Classification (SIC) codes. Designed by the Mexican, United States and Canadian Governments, it was developed to provide a consistent framework for the collection, analysis, and dissemination of industrial statistics.

Its purposes are: (1) to facilitate the collection, tabulation, presentation, and analysis of data relating to establishments, and (2) to promote uniformity and comparability in the presentation and analysis of statistical data describing the North American economy. NAICS is used by Federal statistical agencies that collect or publish data by industry. It is also widely used by State agencies, trade associations, private businesses, and other organizations.

Mexico's Instituto Nacional de Estadística, Geografía e Informática (INEGI), Statistics Canada, and the United States Office of Management and Budget, through its Economic Classification Policy Committee, collaborated on NAICS to make the industry statistics produced by the three countries comparable. NAICS is the first industry classification system developed in accordance with a single principle of aggregation, the principle that producing units that use similar production processes should be grouped together in the classification.

For the three countries, NAICS provides a consistent framework for the collection, tabulation, presentation, and analysis of industry statistics used by government policy analysts, by academics and researchers, by the business community, and by the public. However, because of different national economic and institutional structures as well as limited resources and time for constructing NAICS, its structure was not made entirely comparable at the individual industry level across all three countries. For some sectors and sub sectors, the statistical agencies of the three countries agreed to harmonize NAICS based on sectoral boundaries rather than on a detailed industry structure. NAICS comparability is limited to the sector level for wholesale trade, retail trade, and public administration.

The four principles of NAICS are:

- (1) NAICS is erected on a production-oriented conceptual framework. This means that producing units that use the same or similar production processes are grouped together in NAICS.
- (2) NAICS gives special attention to developing production-oriented classifications for (a) new and emerging industries, (b) service industries in general, and (c) industries engaged in the production of advanced technologies.
- (3) Time series continuity is maintained to the extent possible.
- (4) The system strives for compatibility with the two-digit level of the International Standard Industrial Classification of All Economic Activities (ISIC Rev. 3) of the United Nations.

NAICS uses a hierarchical structure to classify establishments from the broadest level to the most detailed level using the following format:

Sector	2-digit	Sectors represent the highest level of aggregation. There are 20 sectors in NAICS representing broad levels of aggregation.
Sub-sector	3-digit	Sub sectors represent the next, more detailed level of aggregation in NAICS. There are 100 sub sectors in NAICS.
Industry Group	4-digit	Industry groups are more detailed than sub sectors. There are 317 Industry groups in NAICS.
NAICS Industry	5-digit	NAICS industries are the level that, in most cases, represents the lowest level of three country comparability. There are 725 five-digit industries in NAICS.
National Industry	6-digit	National industries are the most detailed level of NAICS. These industries represent the national level detail necessary for economic statistics in an industry classification. There are 1179 U.S. industries in NAICS United States, 2002.

Source: U.S. Census Bureau, NAICS 2002.

## CMO

The dimensions and characteristics of Mexican's geography together with a vast diversity of natural resources make Mexico a country where every kind of economic activity is possible. For these reasons we find a much diversified occupational structure where statistical treatment becomes really complex.

To tackle this problem the Mexican Classification of Occupations (CMO for its name in Spanish) was created. In the CMO the occupations are classified in four different levels of aggregation using the following format:

Main Group	2-digit	Main groups represent the highest level of aggregation. There are 19 main groups in CMO representing broad levels of aggregation.
Subgroup	3-digit	Each main group is formed of subgroups which represent the next, more detailed level of aggregation. There are 137 subgroups in CMO.
Unitary Group	4-digit	Unitary groups are more detailed than subgroups. There are 465 unitary groups in MCO.
Individual Occupation	4-digit	Individual Occupations represents the lowest level of aggregation. There are 9600 in MCO. They are identified by a four-digit id that corresponds to the id of the Unitary Groups in which they belong. For this reason, using this structure it is not possible to identify a level of desegregation beyond unitary groups.

Source: INEGI. Clasificación Mexicana de Ocupaciones (CMO) 1996.

## Appendix E:

### Right answers to Raven's Progressive Matrices Tests

Raven test applied to children between 5 and 12 years old			Raven test applied to adult members between 13 and 65 years old	
Matrix number	Right answer		Matrix number	Right answer
1	8		1	5
2	4		2	2
3	5		3	3
4	1		4	2
5	2		5	3
6	5		6	5
7	6		7	5
8	3		8	6
9	7		9	1
10	8		10	4
11	7		11	3
12	6		12	2
			13	6
			14	2
			15	3
			16	6
			17	3
			18	5



## Appendix F

### Names of Data Files for the Community Survey

Book	File Name	Sections Content	Level of Observation	No. Records
Schools' Questionnaire	esc_am1	AM- Students and teachers (1)	Facility	1,172
	esc_am2	AM- Students and teachers (2)	Facility	1,172
	esc_am3	AM- Students and teachers (3)	Facility	1,171
	esc_ca1	CA- Education costs and school aids (1)	Facility	2,344
	esc_ca2	CA- Education costs and school aids (2)	Facility	9,376
	esc_ca3	CA- Education costs and school aids (3)	Facility	1,172
	esc_contra1	Back page (1)	Facility	7,032
	esc_contra2	Back page (2)	Facility	2,344
	esc_de	DE- School's principal	Facility	1,172
	esc_dg	DG- General descriptions	Facility	1,172
	esc_fe1	FE- School's function (1)	Facility	1,172
	esc_fe2	FE- School's function (2)	Facility	10,548
	esc_in	IN- Infrastructure	Facility	1,172
	esc_ne1	NE- Interview session notes(1)	Facility	1,172
	esc_ne2	NE- Interview session notes(2)	Facility	1,172
	esc_ne3	NE- Interview session notes(3)	Facility	1,172
	esc_ne4	NE- Interview session notes(4)	Facility	1,172
	esc_ne5	NE- Interview session notes(5)	Facility	1,172
	esc_ne6	NE- Interview session notes(6)	Facility	1,172
	esc_ne7	NE- Interview session notes(7)	Facility	1,172
	esc_portad	Cover	Facility	1,172
	esc_prm1	PRM- Questions to teachers(1)	Facility	1,172
	esc_prm2	PRM- Questions to teachers(2)	Facility	1,172
Book of Community Characteristics	loc_ac	AC- Communitarian activities	Facility	150
	loc_as1	AS- Social attendance(1)	Facility	150
	loc_as2	AS- Social attendance(2)	Facility	452
	loc_asc	ASC- Community social aspects	Facility	150
	loc_bc	BC- Community welfare	Facility	150
	loc_cm1	CM- Migrants clubs(1)	Facility	150
	loc_cm2	CM- Migrants clubs(2)	Facility	300
	loc_contra1	Back page(1)	Facility	1,043
	loc_contra2	Back page(2)	Facility	300
	loc_dn	DN- Natural disasters	Facility	1,523
	loc_ef1	EF- Enterprises and factories(1)	Facility	150
	loc_ef2	EF- Enterprises and factories(2)	Facility	518

	loc_el	EL- Availability of electricity	Facility	150
	loc_fa	FA- Water source and sanitation	Facility	150
	loc_hi	HI- Infrastructure history	Facility	1,217
	loc_i	I- Industry	Facility	150
	loc_inf1	INF- Infrastructure(1)	Facility	150
	loc_inf2	INF- Infrastructure(2)	Facility	3,053
	loc_mt	MT- Transportation	Facility	150
	loc_ne	NE- Interview session notes	Facility	150
	loc_oc	OC- Credit opportunities	Facility	911
	loc_od	OD- Direct observation	Facility	150
	loc_pb	PB- Population	Facility	150
	loc_portad	Cover	Facility	172
	loc_sed	SED- History of the presence of schools	Facility	150
	loc_sm1	SM- History of the presence of health facilities(1)	Facility	150
	loc_sm2	SM- History of the presence of health facilities(5)	Facility	1,526
Book of Small Health Providers				
	peq_agc	AGC- General aspects of the town	Facility	1,167
	peq_contra1	Back page(1)	Facility	7,002
	peq_contra2	Back page(2)	Facility	2,334
	peq_gp1	GP- General data of provider(1)	Facility	1,167
	peq_gp2	GP- General data of provider(2)	Facility	8,169
	peq_med1	MED- Medicines(1)	Facility	1,167
	peq_med2	MED- Medicines(2)	Facility	5,844
	peq_med3	MED- Medicines(2)	Facility	30,342
	peq_ms	MS- Health material, equipment and set of instruments	Facility	1,167
	peq_ne	NE- interview session notes	Facility	1,167
	peq_portad	Cover	Facility	1,172
	peq_pos	POS- Problems when offering services	Facility	10,503
	peq_pt	PT- Traditional midwife	Facility	1,167
	peq_sa	SA- Waiting room	Facility	1,167
	peq_srm	SRM- Medical exploration room	Facility	1,167
	peq_sv1	SV- Services offered(1)	Facility	1,167
	peq_sv2	SV- Services offered(2)	Facility	32,676
	peq_sv3	SV- Services offered(3)	Facility	5,835
Book of Prices in the town				
	pre_contra1	Back page(1)	community	996
	pre_contra2	Back page(2)	community	332
	pre_ne1	NE- Interview session notes(1)	community	166
	pre_ne2	NE- Interview session notes(2)	community	166
	pre_ne3	NE- Interview session notes(3)	community	166
	pre_pa1	PA- Prices of agricultural products(1)	community	166
	pre_pa2	PA- Prices of agricultural products(2)	community	1,840
	pre_pm1_1	PM- Prices in the town(1)	community	166
	pre_pm1_2	PM- Prices in the town(2)	community	11,385
	pre_pm2_1	PM- Prices in the town(3)	community	166
	pre_pm2_2	PM- Prices in the town(4)	community	9,767

	pre_pm3_1	PM- Prices in the town(5)	community	166
	pre_pm3_2	PM- Prices in the town(6)	community	9,306
	pre_portad	Cover	community	166
	pre_prf1	PRF- Prices at drugstore(1)	community	166
	pre_prf2	PRF- Prices at drugstore(2)	community	3,881
Book of Health Service				
Infrastructure	inf_ag1	AG- General aspects of the unit(1)	Facility	903
	inf_ag2	AG- General aspects of the unit(2)	Facility	904
	inf_ag3	AG- General aspects of the unit(3)	Facility	17,176
	inf_ag4	AG- General aspects of the unit(4)	Facility	6,328
	inf_contra1	Back page(1)	Facility	5,424
	inf_contra2	Back page(2)	Facility	1,808
	inf_fm1	FM- Drugstore/Medicines(1)	Facility	904
	inf_fm2	FM- Drugstore/Medicines(2)	Facility	23,504
	inf_lab1	LAB- Laboratory for clinical tests(1)	Facility	904
	inf_lab2	LAB- Laboratory for clinical tests (2)	Facility	13,560
	inf_lb	LB- Direct observation of laboratory	Facility	904
	inf_ne1	NE- Interview session notes(1)	Facility	904
	inf_ne2	NE- Interview session notes(2)	Facility	904
	inf_ne3	NE- Interview session notes(3)	Facility	904
	inf_ne4	NE- Interview session notes(4)	Facility	904
	inf_ne5	NE- Interview session notes(5)	Facility	904
	inf_portad	Cover	Facility	918
	inf_ps1	PS- Personnel and services that the unit has(1)	Facility	904
	inf_ps2	PS- Personnel and services that the unit has(2)	Facility	18,080
	inf_rm	RM- Direct observation of exploration room	Facility	904
	inf_rs	RS- Responsible for the services	Facility	4,520
	inf_se	SE- Direct observation of waiting room	Facility	904
	inf_ser1	SER- Activities and services provided by the unit(1)	Facility	904
	inf_ser2	SER- Activities and services provided by the unit(2)	Facility	904
	inf_ser3	SER- Activities and services provided by the unit(3)	Facility	6,309
	inf_ser4	SER- Activities and services provided by the unit(4)	Facility	17,309
	inf_ser5	SER- Activities and services provided by the unit(5)	Facility	9,040
	inf_ser6	SER- Activities and services provided by the unit(6)	Facility	6,328

---

## Appendix G:

### Classification of Medicines of book of community prices

Group	Medicine
Antibiotics	Ampicillin
	Erythromycin
	Trimetropim
	Cloranfenicol
	Penicillin
Medicine against High blood pressure	Captopril
	Nifedipine
Diuretics	Furosemide
Anti-parasites	Metronidazol
	Albendazol
Medicine against diabetes	Tolbutamida
	Glibenclamida
	Clorpropamida
Analgesics	Acetylsalicylic acid
	Paracetamol
	Naproxeno
Medicine against tuberculosis	Etambutol
	Isoniazida
Medicine against malaria	Cloroquina
	Primaquina
Vitamin tables	Iron tablets
	Folic acid tablets
Solutions	Envelopes of oral electrolytes
	Isotonic solution of sodium chloride
	Glucose solution

## Appendix H:

### Linking Of Schools, Clinics and Small Health Providers in MxFLS-1 Community Survey with MxFLS-1 Households

A substantial part of the household survey was designed to obtain detailed information about schools, clinics and small health providers used (or simply known) by the household members. For example, section ED of book 3A collected information about the name and address of elementary, secondary and high school attended by the respondent. Section HS of Book 3B asked for the exact location of every clinic where the respondent has been hospitalized. Although these facilities were generally located in the same community as the respondent, there were some exceptions. The reason for this can be straight forward. Suppose a family resides in the border of a big community, far from all the facilities including schools and hospitals and there is a school (from another community) located nearer than that of the own community. In this case, there is a great probability that the household's children will assist to the school of the other community.

Note that there is no guarantee that the other community is part of the MXFLS-1 sample. However, since we wanted to create a database as rich as possible, we collected information about schools, clinics and small health providers mentioned by the household survey even if they were located in non-MxFLS-1 communities. Similarly, there were schools, clinics and small health providers located in MxFLS-1 communities that were not mentioned in the household survey. For the same reason, we decided to collect information about them.

It is important to mention that not all the schools, clinics and small health providers were interviewed for the community survey. A representative sample was selected. The next table shows the exact relation between the number of schools, clinics and small health providers in the household and the community survey:

#### LINKING OF COMMUNITY INFRASTRUCTURE WITH MXFLS-1 HOUSEHOLDS

	FACILITIES IN MXFLS-1 COMMUNITIES		FACILITIES IN NON-MXFLS COMMUNITIES		
	<i>Related with a MxFLS-1 household</i>	<i>Non-related with a MxFLS-1 Household</i>	<i>Related with a MxFLS household</i>	<i>Non-related with a MxFLS household</i>	<i>TOTAL</i>
<i>number of schools</i>	970	202	0		1,172
<i>number of clinics</i>	461	150	185		796
<i>number of small health providers</i>	715	201	185		1,101

## Glossary

### A–F

<i>Alianza para el Campo</i>	Government Program to support the revival and development of the Agricultural Sector in Mexico.
Book	Major section of an MxFLS-1 questionnaire (e.g. book C).
CAFE	Computer-Assisted Field Editing, a system used for the first round of data entry in the field, using laptop computers and software that performed some range and consistency checks. Inconsistencies were resolved with interviewers, who were sent back to respondents if necessary.
<i>Carrera Magisterial</i>	Training program for teachers in basic education (kinder garden, elementary and secondary) which consist of five different levels and provides monetary incentives according to the level achieved.
CBTA	<i>Centro de Bachillerato Tecnológico Agropecuario</i> (Technical Farming High School)
CBTIS	<i>Centro de Bachillerato Tecnológico, Industrial y de Servicios</i> (Technical Industrial and Services High School)
CECATI	<i>Centro de Capacitación para el Trabajo Industrial</i> (Center for Qualification in Industrial Work)
CECYT	<i>Centro de Estudiantes de Ciencia y Tecnología</i> (Center of Science and Technology Students)
CIDE	Centro de Investigación y Docencia Económicas.
CMO	Mexican Classification of Occupation
COBACH	<i>Colegio de Bachilleres</i> (Technical High school)
<i>Crédito a la Palabra</i>	Government Program supporting farming of basic grains and lines of production and commercialization through direct loans to farmers.
CTA	<i>Centro Tecnológico Agrícola</i> (Agriculture Technological Center)
Data file	File of related MxFLS-1 variables. For household data, usually linked with only one household questionnaire section.
DICONSA	State owned company aimed to distribute basic products with high quality and competitive prices in benefit of marginalized areas.
DIF	Institute for Integral Family Development
ENADID	Encuesta Nacional de Dinámica Demográfica (National Demographics Survey)
ENE	Encuesta Nacional de Empleo (National Labor Survey)
ENIGH	Encuesta Nacional de Ingresos y Gastos de los Hogares (National Household Income and Expenditure Survey)
<i>Estrato</i>	Variable that displays the classification of communities, in which each household is located, according to their level of urbanization. <ul style="list-style-type: none"> <li>• 1 = for households located in communities with a population of more than 100,000 inhabitants.</li> <li>• 2 = for households located in communities with a population between 15,000 and 100,000 inhabitants.</li> <li>• 3 = for households located in communities with a population between 2,500 and 15,000 inhabitants.</li> <li>• 4 = for households located in communities with a population below 2,500 inhabitants.</li> </ul>
FOLIO	Household identifier. In data bases the variable is “folio”.

### G–K

Household	A person or group of people, related or unrelated by biological bonds, who usually live together in a part of or in an entire building/dwelling and usually consume meals prepared with a common budget on the same stove/oven and even use the same tools for preparing the meals.
Household member	Any person who usually lives in the household, regardless of his presence or temporary absence. For example someone on vacation or who has left the household temporarily (for less than one year) for labor reasons is considered a household member. A person who has lived in the household for one year or more or who has lived in the household for less than one

year but is planning to stay in the household for a year or more is considered a household member. The guests who fulfill the criteria mentioned above and who sleep in the household, share the meals prepared by the household and are free to use the kitchen. Domestic servants or any other household workers who fulfill the criteria mentioned above are considered household members.

IMSS	<i>Instituto Mexicano del Seguro Social</i> (Mexican Social Security Institute)
ISSSTE	<i>Instituto de Seguridad y Servicios Sociales de los Trabajadores del Estado</i> (Institute of Security and Social Services for Public Workers)
INEGI	<i>Instituto Nacional de Estadística, Geografía e Informática</i> (Mexican National Bureau of Statistics)
INPER	<i>Instituto Nacional de Perinatología</i> (National Institute of Perinatology)

#### **L–O**

Look Ups (LU)	Process of manually checking the paper questionnaire against a computer-generated set of error messages produced by various consistency checks. LU specialists had to provide a response to each error message; often they corrected the data.
LS	Individual identifier. In data bases the variable is “ls”.
Main respondent	An MxFLS-1 respondent who answered an individual book (3a, 3b, 4 or 5).
MxFLS-1	Mexican Family Life Survey.
Municipality	An urban district generally composed of a clearly defined territory and self-government.
Section	Topical subsection within an MxFLS-1 survey questionnaire <i>book</i> .
NAICS	North American Industrial Classification System
Oportunidades	Social program consisting of monetary transfers created to reduce poverty from an integral perspective.
“Other” responses	Responses that did not fit specified categories in the questionnaire.

#### **P–R**

PEMEX	<i>Petróleos Mexicanos</i> (Mexican Oil Company)
Procampo	Direct subsidy provided by the federal government to farmers according to the size of their plots of land.
Proxy (book)	The Proxy Book was designed to collect information on members of the household that were not able to answer the individual questionnaires (they were either away or very sick) so a different member of the household (generally the household head or spouse) reported the information.
Raven’s test	Raven’s Progressive Matrices Test. Tests applied to household members between 5 and 65 years old to measure their cognitive ability.
Rural community	Community with population below 2,500 inhabitants.

#### **S–Z**

<i>Secuencia</i>	Name of the variable that identifies the particular observation level when the data are arranged in grids and the level of observation is something other than the household or individual.
SEMAR	<i>Secretaría de Marina</i> (Mexican Ministry of Navy )
SEDENA	<i>Secretaría de Defensa Nacional</i> (Mexican Ministry of Defence)
SEDESOL	<i>Secretaría de Desarrollo Social</i> (Mexican Ministry for Development)
SEP	<i>Secretaría de Educación Pública</i> (Mexican Ministry of Education)
Special Codes	Codes 5, 6, 7, 8, 9 or multiple digits beginning with 9. Special codes were entered by the interviewer to indicate that numeric data are missing because response was out of range, questionable, or not applicable; or respondent refused to answer or did not know.
SSA	<i>Secretaría de Salud</i> (Mexican Ministry of Health)
System Missing Data	Data properly absent because of skip patterns in the questionnaire.
UIA	Universidad Iberoamericana.
Urban community	Community with population above 2, 500 inhabitants.

