User´s Guide: Mexican Family Life Survey 2005

LUIS RUBALCAVA
División de Economia

Centro de Investigación y Docencia Económicas A.C.

GRACIELA TERUEL
Departamento de Economía

Universidad Iberoamericana

October, 2007

Preface

This document describes the basic principles found in MxFLS-2 allowing analysis and access to the information contained.

A primary goal of the Mexican Family Life Survey (MxFLS) is to create a longitudinal and multi-thematic database. On the one hand, this allows a long term tracking of individuals regardless of change in residence, newly formed homes or home conversions; and on the other hand, these characteristics will be beneficial in their use of a single tool collecting economic, demographic and health population indicators of the Mexican population. The survey's first round (MxFLS-1) took place during the year 2002 reaching a sample size of 35 thousand individual interviews in 150 urban and rural locations throughout the country. The second round (MxFLS-2) ended during 2006 with 90 per cent a re-contact rate at the household level. During this second operational phase of tracking and interviewing original households between 2002 and 2006, those who migrated to the United States were contacted at a rate of over 91 per cent. Within this content, MxFLS-2 provides statistical evidence on the flow of internal migration into the United States, on population and on the welfare dynamics of the Mexican people that relocate within the country and in the US. This is regardless of whether or not they remain or return to their national territory and/or community of origin.

MxFLS-2 has a sample size of approximately 40 thousand individuals due to the unfolding and expansions associated with the interviewers' cycle of life. The third round (MxFLS-3) is scheduled for 2009.

MxFLS-1 and MxFLS-2 databases and documents are public domain and can be accessed at: www.ennvih-mxfls.org

MxFLS-2 was a result of a collaboration with joint efforts from researchers of the University Iberoamericana (UIA), the Center of Economic Research and Teaching (CIDE: *Centro de Investigación y Docencia Económicas*), the National Institute of Public Health (INSP: *Instituto Nacional de Salud Pública*), and the University of California, Los Angeles (UCLA).

MxFLS-2 information, documents and other files will be of interest to the scientific community, to the social political debate in Mexico, and to comparative research of other countries.

MxFLS-2 Documents:

1. Rubalcava, Luis and Teruel, Graciela (2006). "Mexican Family Life Survey, Second Round", Working Paper, www.ennvih-mxfls.org. When using data from MxFLS-2, we ask that you cite this as your reference.

   1.1. Project Design
   1.2. Objective, field work and response rates for the interview, with an emphasis on MxFLS-2.
   1.3. MxFLS-1 and MxFLS-2 Results

2. Berumen (2007). "Sample Design". Description of MxFLS-2 samples. Working Paper, University Iberoamericana. Mexico City.

3. Second round of Household questions for the Mexican Family Life Survey. Spanish and English versions available.

4. Household survey Code manuals for the second round of the Mexican Family Life Survey. Spanish and English versions available.

    4.1. Variable descriptions of the MxFLS-2 questionnaires and their location within the data files.

5. Second round of the Community questionnaires for the Mexican Family Life Survey. Spanish and English versions available.

6. Second round of the Community Survey Code manuals for the Mexican Family Life Survey. Spanish and English versions available.

    6.1. Variable descriptions of the MxFLS-2 questionnaires and their location within the data files.

7. Users Guide for MxFLS-2.

    7.1. Structure description and data formats for MxFLS files.
    7.2. Guidelines for data use, with emphasis on a cross and joint use of MxFLS-1 and MxFLS-2.

ACKNOWLEDGMENTS

iii

Content

1. Introduction

The main objective of MxFLS is to create a longitudinal and multi purpose database with a minimum time span of 10 years. This database will allow for a better understanding of economic, demographic and social transitions that the Mexican people encounter during the length of time this project remains active.

In 2002 researchers from UIA and CIDE, in collaboration with INEGI, carried out the first round of the survey (MxFLS-1). Designated as the base line, this first round of surveys collected detailed information on the socioeconomic, demographic and health status of all household members on a sample size of approximately 8,400 homes located throughout the country. The survey is representative of national, urban-rural and regional levels[1].

On an individual level, MxFLS-1 collected detailed information on each household member including: level of education, retrospective migration background, marriage, fertility and any victimization occurring in the household (robbery, theft, kidnaps); work force participation; adult work earnings; money transfers and in kind; time allocation for adults and children; credit and loans; investments in human capital and the decision making process; health status perceptions, objective measures pertaining to their state of health (weight, size, waist and hip measurements, blood pressure levels, standard blood lab tests); reproductive health of all women within fertile age range and contraceptive methods and usage. All individuals within their appropriate age range were subject to Raven's Progressive Matrices, with the purpose of including a measure of the candidate's cognitive state.

Given that a person's welfare also depends on their environment (quantity and quality of the infrastructure and services available within their particular province/region), MxFLS-1 carried out a parallel fieldwork survey interviewing health centers, schools, community leaders and commercial establishments.

This was conducted with the goal of capturing qualitative and quantitive data on a provincial/ geographical location level regarding infrastructure and services provided at schools, and the current and past physical infrastructures and socioeconomic status of health centers. These survey tools where applied at all geographical locations where the interviewee's live. The following information was collected: current and past economic and physical community infrastructure (for example, salaries, public services available, transportation and communication infrastructure, natural disasters, crime, school and health clinic infrastructures, services provided at health clinics, basic product prices, and others).

The survey's panel design allows a continuous tracking of those individuals and those families interviewed at the base line, regardless of their geographic location or economic mobility. Continuing with the MxFLS panel form will considerably enrich the information available and allow us to fully analyze the welfare and associated changes throughout time within our country.

This is why researchers from the Center of Economic Research and Teaching (CIDE) and from the University Iberoamericana (UIA) united their endeavors to carry out the second round of MxFLS surveys. This second round (MxFLS-2) began in mid 2005 and ended in 2006. As part of the panel structure, the fundamental purpose of MxFLS-2 was to re-interview all individuals and households who were originally

---

1 Based on the regional definition of the *Plan Nacional de Desarrollo*, Mexico. 2000-2006.

interviewed during the base line (MxFLS-1) and all additional individuals and households who grew out and developed from our original sample. Given the importance of the migration phenomenon experienced in the country, this was one of the most challenging and important tasks of the project. In spite of this, a re-contact rate at the household level reached 90% of our original sample size.

Another vast challenge was presented in searching, tracking and interviewing those individuals who migrated to the United States during the rounds. The perseverance of the MxFLS team in conjunction with support received from UCLA, were key elements in reaching the re-contact rate of immigrants in the Unites States of over 91%.

Prior to MxFLS-2, there was no source of information in Mexico that would present detailed information on individual and family conditions of migrating populations before and after their journey. MxFLS-2 provides gathered information from both sides of the border providing additional details on the migration dynamics of Mexico-United States. With public policy involvement, there could be a better understanding of the complex relationship between Mexican emigrants and their families of origin. This will be transcendental in the welfare of the population.

MxFLS-2 follows the content, design and structure of the MxFLS-1 household and community questionnaires. However, in addition to the questions established in the base line, MxFLS-2 incorporates new sections that contain questions relating to individual expectations, seasonal preferences, altruism and risk taking. In addition, there are also questions regarding paternity and child care.

1.1 Institutional Collaboration

Researchers form the Center of Economic Research and Teaching (CIDE) and from the University Iberoamericana (UIA) united their efforts in carrying-out the second round of MxFLS surveys (MxFLS-2). Investigators from diverse institutions participated in the design of new modules and questions integrated into MxFLS-2. This was especially accomplished through the specialized assistance of experts at the University of California, Los Angeles departments of Economics and Sociology (UCLA).

During MxFLS-2 fieldwork, collective agreements were reached with two prestigious institutions. *Berumen and Associates*, a company who specializes in conducting household surveys, was responsible for the operational component of collecting household information. In turn, Mexico's *National Institute of Public Health* (INSP) had personalized personnel who where responsible for collecting anthropometric measure pertaining to health, the design of diet and health questionnaires and the collection of data at the community level.

1.2 Project Planning

The planning and administration of the project is under the direction of Luis Rubalcava, Graciela Teruel, Duncan Thomas and Ellizabeth Frankenberg.

1.3 MxFLS-2 Sponsors

Prior activities associated with conducting the survey, collection and publication of data, where financed by diverse institutions. Their contributions were essential in order to complete the second round of MxFLS. The MxFLS-2 team would like to thank the following institutions:

- Banamex

- DHL
- CONACYT-SEDESOL Funds
- Mexico's National Health Care System (Instituto Mexicano del Seguro Social-IMSS)
- The United States National Institute of Child, Health and Development (NICHD)
- Mexico's State Department of Social Development (Secretaria de Desarrollo Social -SEDESOL)
- Mexico's State Department of Health (Secretaría de Salud)
- The University of California Institute for Mexico and the United States, (UC MEXUS)
- University Iberoamericana, (UIA, FICSAC)

2. MxFLS-2 Design

2.1 MxFLS-2 Questionnaire Design

MxFLS-2 follows the content, design and structure of MxFLS-1 questionnaires. Additionally, MxFLS-2 contains innovative modules collecting data on child upbringing, individual expectations, seasonal preferences, altruistic behavior and risk taking. This will allow, for example, a study on the relationship between individual expectations and preferences, and the behavior among household members in the decision making process within their household and community environment. UCLA, UIA and CIDE specialists implemented the design of these new modules.

As well as the 3 household books: C (Control or Log Book), I (Household Consumption), II (Household Economy) and the 8 individual books: IIIA and IIIB (Characteristics of Household Members who are 15 years old or older), EA (Cognizant State of Household Members between the ages of 13 and 65), EN (Cognizant State of Household Members between the ages of 5 and 12), Proxy book (Absent adults) and Book S (Health) that compose the MxFLS-1 design, MxFLS-2 incorporates a new book; Book D (Diet Book). The objective of this new book is to identify the diet patterns of household members. Its design was created by Dr. Juan Rivera Dommarco from INSP.

All additional MxFLS-1 modules and questions were tested at distinct urban, rural and indigenous population zones.

Once the utilization of these new questions was accepted, they were implemented to the MxFLS questionnaire and pilot trial runs took place for the final books. In the state of San Luis Potosi, just as with MxFLS-1, trained interviewers along with specialized health personnel carried out a general pilot test run on the survey protocol of information, emigrant search, the coordination of interviewers with health personnel, cleaning fieldwork data and all activities associated with fieldwork.

2.2 Household and community information

In a single database, MxFLS-2 collects detailed information on the inhabited household and community. The household information includes:

- Household and Self consumption
- Decisions taken regarding savings, assets, debts, employment and non-employment earnings, public/private transfers, and household/individual social program participation.
- Retrospective information on education, school interruptions and grade repetition.
- Retrospective information on work related decisions for all household members, including children.
- Time allocation
- Demographic information of non-resident, household family members
- Temporary and Permanent migration information
- Individual health status and chronic illness information
- Individual mental health status information
- Information on anthropometric measures for all household members (weight, size, waist and hip measurements, blood pressure, hemoglobin, blood sugar, cholesterol, and C-reactive protein levels)
- Detailed retrospective information on the use of health services
- Pregnancy history, contraceptive usage and methods, post-natal health services
- Marital history and decisions made within the household
- Crime incidents at the household and individual levels
- Detailed information on the family business and use of land
- Information on risk-taking situations and on the decision making process
- Information regarding a trustworthy environment
- Preferred time allocation information
- Child upbringing information
- Household diet information (representational information taken on behalf of mother and child household members)

Parallel to the household survey information, MxFLS-2 collected data at the community level. As with the household questionnaires, the design, structure and content of the community questionnaires are similar to the MxFLS-1 community questionnaires. The objective of the MxFLS-2 community questionnaires is to complement the household information with quantitative and qualitative data on the location infrastructure and local household prices with socio-demographic characteristics. During MxFLS-2, the only book information collected was that on infrastructure and prices.

The community level information includes:
- General population information of that particular location.
- Social aspects within the location

- Natural disaster information
- Infrastructure history
- Social service assistance
- Educational services
- Medical Services
- Community activities
- Methods of transportation
- Sources of water
- Electricity
- Credit opportunities
- Industries, companies and factories
- Immigrant clubs
- Prices of goods and services at various commercial establishments within the particular location.

2.3 Confidentiality

In order to insure confidentiality of MxFLS-2 participants, all identifying information has been eliminated from public files. This information includes: names, address, previous location information, school names and addresses, health care centers and private health care providers that assist household members, and others.

In the same fashion, in order to ensure confidentiality of services within the location, the MxFLS team replaced the INEGI location codes with an internal coding system.

3. MxFLS-2: Elements derived from MxFLS-1

The elements derived from MxFLS-1 are presented in this section.

3.1 Re-interviews: MxFLS-1 Households and Individuals

The purpose of MxFLS-2 is to re-interview all individuals and households who were interviewed during the base line (MxFLS-1) and all individuals and households who, due to the growth and family developments from our original sample size, were added on. For each of the original households (Home Panel), a pre-printed list was created with the most relevant socio-demographic characteristics of 2002 household members, including: name, age, date of birth, sex and relationship to the head of household. The pre-printed list was of great use during the re-interview process because it indicated the specific people who had to be interviewed or the specific people the interviewer had to ask for.

Due to the important migration phenomenon in Mexico, one of the most important challenges during MxFLS-2 was to locate the households and individuals who changed residence between 2002 and 2005. From the original household and individuals surveyed in 2002, **6.88% and 11.73%** of them respectively, no longer

had the same address. The responsibility of tracking households and individuals implied the development of new and complex search strategies. In cases where the Home Panel was no longer at the original address, the interviewers would conduct investigations and search for the final destination of that particular household. In order to do so, the interviewers had additional information indicating the possibilities of locating that household. During MxFLS-1, a section in book C was designed (a re-contact section) where in the case of the household were to move, the address and telephone number of a close relative could assist in providing current household information. A re-contact directory was created with this information and used to identify the current household location.

3.2. Re- contact and tracking of MxFLS-2 households and individuals

A series of re-contact operatives with previously established protocols took place in order to locate homes and individuals, such as verifying with neighbors and community members if those specific people did indeed live in the directories referred address list. During the visit of the original household address, there was an attempt to collect information on the new location of the home from neighbors and key informants within the community.

Immigrant tracking was carried out with three strategies: i) Re-contact homes and individuals who moved within the 16 states that make-up the MxFLS sample size, ii) re-contact homes and individuals who moved to states who do not make-up the MxFLS sample size and iii) re-contact individuals and households who migrated to the United states. Special teams were created to collect information about the household diet (by taking representational information from the mother and child household members).

3.3   Home Panels and New Home Classifications

All longitudinal surveys face attrition problems that could stem from, on the one hand, a no-response and on the other, socio-demographic changes that affect the original sample: deaths, births, marriages, separations etc. The most common ways to confront the panel survey attritions are to 1) artificially update the sample size ("refresh" the sample). That is, add households and/or individuals to the original sample who were not necessarily involved with the original household surveys and 2) "naturally" update the sample by adding those individuals who by birth or by unions are now considered members of the home.

In the case of MxFLS, it was decided to refresh our sample size naturally. In this way, a home or individual is defined as panel if they belong to the original sample size and if they were interviewed during 2002. The structure of the home panel could be modified due to a member or members leaving the home, due to deaths or to the addition of new household members. Whatever the case may be, the most interesting variable here are the individuals. That is to say, that it would be necessary to search and interview those home panel members who were interviewed during 2002 and who are no longer home members in 2005. If on the other hand, new individuals who were integrated into the home and were not interviewed in 2002, but are now considered family members, they would be classified as new members and interviewed for the first time.

A new home surges when one or various individuals who were interviewed during 2002 left the original household to form a new home. For example, a person who belonged to an MxFLS home in 2002 and left the household to join another person (whether this person was an MxFLS home member or not) this union comprises a

new home and all people who make up this home, as well as the original home panel member will be interviewed. In this way, we will gain new households and individuals who update our original sample size and resolve the attrition problems we may come across.

All new homes must always be composed of at least one panel member. In a panel home there may or may not be new members.

3.4 Pre-printed lists

With the objectives of verifying information obtained during the first round of surveys (MxFLS-1), correcting possible errors and updating home information in the most effective way, a pre-printed list and informant of home members was provided to the interviewer.

Pre-printed lists are created for the following questionnaire sections:

- **Home members Pre-printed list (Book C)**

Logs all names and personal information (sex, age, date of birth, relationship to head of household, list of survey books answered during the first round) of each and every home member during 2002.

- **Pre-printed list of Centers, Health Facilities and Schools attended by home members (Book C).**

Records the health centers, private health care providers and schools attended by household members during 2002 and who comprised part of the community sample of MxFLS-1.

- **Pre-printed list information pertaining to children, siblings, and other people who are not household residents (Book IIIB).**

Logs information on children, siblings and people who are not household residents (corresponding to the transfer section in Book IIIB), and information provided by home members during 2002.

Pre printed lists are only used for panel homes.

4. MxFLS-2 Structure

MxFLS-2 follows the same structure as MxFLS-1. Generally, the same books and the same questions were applied, however, some questions were added and some were eliminated. The following table shows the book structure during MxFLS-1 and MxFLS-2 as well as the appropriate informant for each book. The only difference as far as the books are concerned, is the addition of the MxFLS-2 Diet book, which did not exist in 2002. The person responsible for meal preparations, and a child household member between the ages of 5 and 11 are the ones who answer this new survey book. As far as the appropriate informants are concerned, they would be the same panel members who answered in 2002.

Table 1. Book Structure for homes and appropriate informant for MxFLS-1 and MxFLS-2

MxFLS-1 Book	MxFLS-1 Informant	MxFLS-2 Book	MxFLS-2 Informant
Home			
C	HM* 18 years old or older (Head of household preferably)	C	Same as in MxFLS-1
I	HM 18 years old or older (preferably the person responsible for meals/household expenses	I	Same as in MxFLS-1
II	HM 18 years old or older (Head of household preferably)	II	Same as in MxFLS-1
Individual			
IIIA	All HM's 15 years old or older	IIIA	Same as in MxFLS-1
IIIB	All HM's 15 years old or older	IIIB	Same as in MxFLS-1
IV	All female HM's between the ages of 14 and 49	IV	Same as in MxFLS-1
V	All HM children under the age of 15	V	Same as in MxFLS-1
EA	All HM's between the ages of 13 and 65	EA	Same as in MxFLS-1
EN	All HM's between the ages of 5 and 12	EN	Same as in MxFLS-1
PROXY	A HM 18 years old or older who would know details about the absent household member (1 book for every absent member)	PROXY	Same as in MxFLS-1
HEALTH	All HM's	HEALTH	Same as in MxFLS-1
-		DIET	HM responsible for preparing meals and a HM child between the ages of 5 and 11.

* HM: Household Member

The following table shows the difference between MxFLS-1 and MxFLS-2, in terms of questions and questionnaire sections. The main difference between the two rounds of surveys are focused on questions regarding risks and trust, which are found in book IIIB and book II, and in regards to the upbringing of children found in book V. Some additional questions were added in book C section LS as a tracking technique and as a means to classify individuals, new homes and panel homes.

Table 2. Difference between questions and sections in MxFLS-1 and MxFLS-2

Book	MxFLS-1 Sections	MxFLS-2 Sections	Differences between MxFLS-1 and MxFLS-2
C	DF (Definition)	DF (Definition)	Question DF02 is added to distinguish between a new and panel home. A pre-printed list for section LS book C must be used if this is a panel home.
	LS (List of Household Members)	LS (List of Household Members)	Questions LS001-LS004 are added to ensure that all household members are listed. Questions LS19-LS19f are added to verify information on absent household members. The Address List section is added.
	CV (Household Characteristics)	CV (Household Characteristics)	None
	RC (Re-contact Information)	RC (Re-contact Information)	None
	SP (Health and Family Planning)	SP (Health and Family Planning)	None
	EH (HM's Education)	EH (HM's Education)	None
	CVO (Direct Observations of household characteristics)	CVO (Direct Observations of household characteristics)	None
	NE (Interview Section Notes)	NE (Interview Section Notes)	None
I	CS (Food Consumption)	CS (Food Consumption)	None
	NE (Interview Section Notes)	NE (Interview Section Notes)	None
II	SU (floors)	SU (floors)	None
	INR (Rural Earnings)	INR (Rural Earnings)	None
	NNA (Non-agricultural business)	NNA (Non-agricultural business)	None
	AH (Household Assets)	AH (Household Assets)	None
	CRH (Household Credit)	CRH (Household Credit)	None
	IN (Non-employment household income)	IN (Non-employment household income)	None
	SE (Household economic shocks)	SE (Household economic shocks)	None
	VLH (Household violence and victimization)	VLH (Household violence and victimization)	K through X were added to the VLH questions capturing trust and altruism matters.
	NE (Interview Section Notes)	NE (Interview Section Notes)	None

IIIA	ED (Education)	ED (Education)	None
	IE (School Interruptions)	IE (School Interruptions)	None
	AH (Household Assets)	AH (Household Assets)	None
	IIN (Non-Employment Individual Income)	IIN (Non-Employment Individual Income)	None
	HM (Marriage history)	HM (Marriage history)	None
	DH (Household decisions)	DH (Household decisions)	None
	MG (Permanent migration)	MG (Permanent migration)	None
	MT (Temporary Migration)	MT (Temporary Migration)	None
	TB (Work)	TB (Work)	None
	ATA (Time allocation)	ATA (Time allocation)	None
	SHI (Individual Shocks)	SHI (Individual Shocks)	Additional expectations questions were added.
	VLI (Household violence and victimization)	VLI (Household violence and victimization)	None
	NE (Interview Section Notes)	NE (Interview Section Notes)	None
IIIB	GH (Habits and preferences)		None
	-	EE (Expectations)	Sections EE was created (Expectations)
	-	RG (Risks)	New MxFLS-2 section
	-	CO (Trust)	New MxFLS-2 section
	ES (State of Health)	ES (State of Health)	None
	SM (State of Mind/Mood)	SM (State of Mind/Mood)	None
	EC (Chronic Illnesses)	EC (Chronic Illnesses)	None
	ATS (Self Help)	ATS (Self Help)	None
	CE (External Consultation)	CE (External Consultation)	None
	HS (Use of Hospitalization services)	HS (Use of Hospitalization services)	None
	CA (Insurance)	CA (Insurance)	None
	RE (Family Contacts in the United States)	RE (Family Contacts in the United States)	None
	CR (Credit)	CR (Credit)	None
	PR (Preferences)	PR (Preferences)	None

	-	FH (Child Upbringing)	New MxFLS-2 section
	TP (Non-resident Parent Transfer)	TP (Non-resident Parent Transfer)	None
	TH (Non-resident sibling transfers)	TH (Non-resident sibling transfers)	None
	THI (Non-resident child transfers)	THI (Non-resident child transfers)	None
	TO (Non-resident transfer of other people)	TO (Non-resident transfer of other people)	None
	NE (Interview Section Notes)	NE (Interview Section Notes)	None
IV	RES (Pregnancy Summary)	RES (Pregnancy Summary)	None
	HE (Pregnancy History)	HE (Pregnancy History)	None
	AC (Contraception)	AC (Contraception)	None
	NE (Interview Section Notes)	NE (Interview Section Notes)	None
V	EDN (Child Education)	EDN (Child Education)	None
	EMN (Child Employment)	EMN (Child Employment)	None
	ATN (Child Time- Assignation)	ATN (Child Time- Assignation)	None
	CEN (Use of external child consultations)	CEN (Use of external child consultations)	None
	ESN (Child state of health)	ESN (Child state of health)	None
	HSN (Use of child hospitalization services)	HSN (Use of child hospitalization services)	None
	AUTN (Child self-treatment)	AUTN (Child self-treatment)	None
	VAC (Child Vaccination)	VAC (Child Vaccination)	None
	-	FH(Child upbringing)	New MxFLS-2 section
	NE (Interview Section Notes)	NE (Interview Section Notes)	None
S	SA (Health Measures)		Levels of Blood Sugar, Cholesterol and C-reactive proteins are added.
	NE (Interview Section Notes)		
D	-	CA (Food consumption)	New MxFLS-2 section
	-	NE (Interview Section Notes)	New MxFLS-2 section
EA	ECA	ECA	None

	NE (Interview Section Notes)	NE (Interview Section Notes)	None
EN	ECN	ECN	None
	NE (Interview Section Notes)	NE (Interview Section Notes)	None

4.1 Retrospective Information

There are some modules that were designed to collect retrospective information on household members from books IIIA, IIIB and IV. Examples of these modules are education, marriage history, migration and participation in the work force. The individuals who answered these modules during MxFLS-1 (a panel member) did not have to re-answer the complete history of these modules, only update details as of 2002 for the MxFLS-2 interview. The individual classification as a panel or new member is found accordingly in the interviewer's pre-printed list based on the 2002 books.

Generally, MxFLS-2 collected less information if the case was panel as opposed to new. The questionnaires were structured as 1) to collect the same retrospective information of new and panel members and 2) an update of information collected during MxFLS-1 for panel members, that is, questions are asked regarding specifics as of 2002. This way we have complete retrospective information for panel individuals of MxFLS-2.

Table 4 gives a summary of the differences between the data collected for panel and new individuals in retrospective modules, and their implications by creating a complete panel history.

Table 3. MxFLS-1 and MxFLS-2 Retrospective Information

Section	New Members	Panel Members	Complete Panel Members History
IIIA-ED (Education)	Last level of education attained	Last level of education attained as of the 2001/2002 school year.	Use the MxFLS-1 education module information for schooling prior to 2002. Schooling between 2002 and 2005 is reported in MxFLS-1 and MxFLS-2
IIIA-ED (Repeated Grades)	Repeated school-grade level as of elementary school	School-grade level repeated as of 2002	Use the MxFLS-1 education module information for schooling prior to 2002. Schooling between 2002 and 2005 is reported in MxFLS-1 and MxFLS-2
IIIA-HM (Marriage History)	Complete Marital History	Marriage History as of 2002	For all interviewees who have not married prior to 2001, MxFLS-2 gives a complete marital history. Details on marriages that ended prior to 2001 are found in MxFLS-2.
IIIA-MG (Permanent Migration)	All questions relating to birth (place of birth and all relocations as of the age of 12)	All relocations as of the age of 12	Use MxFLS-1 for place of birth and place of residence as of the age of 12
IIIA-TB (Work) Verify TB54x	Questions related to most recent employment, employment during 2003 and first employment	Questions related to most recent employment, employment during 2003	Additional panel interview information can be found in MxFLS-1

4.2. MxFLS-2 service and community re-interviews

As previously mentioned, one of the main MxFLS-2 objectives is to become aware of factors that determine the individual welfare and it's evolution throughout time. In this way, MxFLS-2 aspires to re-interview all homes, individuals, locations and commercial establishments who were interviewed during MxFLS-1.

MxFLS-2 contains information comprised of survey participants on 300 locations throughout the country.

5. File Structure

To analyze with ease, this section contains information on the organizations, names and other specific file details. Additional information on the database can be found in the questionnaires and codification manuals.

For a better analysis, every home and community questionnaire page includes the file names that contains the information found on that page. There is a codification manual for each questionnaire that describes the data file content for each, as well as the level of observation represented.

5.1. Basic File Organization

The household and community data files are available in STATA format.

The MxFLS-2 data files are organized in three main groups:

- Detailed documentation of the Mexican Family Life Survey, second round (MxFLS-2)
- Household survey data files
- Community survey data files

5.2. Household Surveys

The household survey files include the database, the questionnaires and codification manuals. The database is available in Stata 8.0 format, and for those who are interested in working with a different format, Stata Transfer software is recommended. This software can convert the files from Stata to ASCII, dBase, Excel, SAS, SPSS, Statistica and SYSTAT.

Household survey data files:

- Every zipped file contains 12 household survey books. As a reference, the letters hh indicate the home; numbers 05 indicate 2005 and dta indicates Stata format for those data files.
- The household survey is comprised of twelve interview books. The files are described in the following manner: hh05dta_byy, one file for every

module of each household book; where dta indicates the corresponding data files and yy is equivalent to the name of the book (c, 1, 2, 3A, 3B, 4, 5, S, D, EA, EN and X for the Proxy book).

- There are corresponding questionnaire files and codification manuals for all household survey data files.

It is easy to understand the file nomenclature. The first characters are associated with the specific book questionnaire followed by the specific module characters and the final number denotes any multiple module database files.

xxx_xxx	X
Book_Module	File sequence
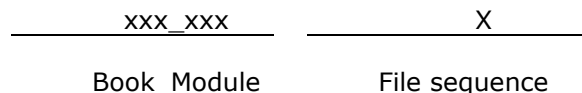
For example, in regards to the permanent migration module (MG) in book 3A, the first database: iiia_mg means that the file contains information from book 3A, module MG. The following: iiia_mg1 denotes a second file of information from book 3A, module MG.

5.2.1 Level of observation and Identifiers

The survey data has been organized in the best possible way, so that the level of observation contained in each file is at the household or individual level. If the level of observation is at the household level, the folio variable identifies only one observation. If it is at the individual level, the folio and Ls are required in order to identify a person.

In MxFLS-2, the folio is a variable made up of 8 digits in the following way:

x x x x x x	x x
Folio 2002	original / new
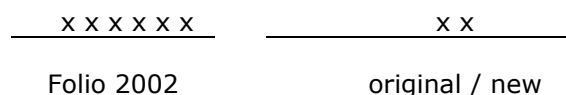
The last two digits, 00 are designated for original homes during MxFLS-1 that do not move. For new homes (an individual who has moved and created a new home), the first 6 digits are the same as the individual's original folio number from where he or she came from, and the last two digits are made up of the individual's designated ls number used during MxFLS-1. The MxFLS-2 Ls number is the same as the MxFLS-1 for individuals who remain in their panel homes. If there are additional household members, the Ls number is consecutive. In other words, if there were three members during MxFLS-1 and the Ls was 01, 02 and 03, the new member during MxFLS-2 would be 04 regardless if one of the previous members has left the home or is deceased.

The new household individuals who were original (panel) home members will not be assigned to the same Ls as in MxFLS-1.

For example, a home folio "00001000" during MxFLS-1 which had four members contained the following Log Book information:

Folio 2005	Ls
00001000	01
00001000	02
00001000	03
00001000	04

But during MxFS-2, the member with Ls "03" has left the original home and formed his own home with two additional members. For this new home the folio was created the following way:

Folio 2005	Ls
00001003	01
00001003	02

Similarly, a new variable called pid_link was created during MxFLS-2. This is a unique individual indicator that stays constant throughout the different survey rounds allowing us to paste individual information throughout the various rounds. Attaching the folio to the individual's MxFLS-1 Ls created the pid_link, whether or not the particular individual remains in this household. For new individuals who were surveyed during this round, the pid_link was comprised of joining the MxFLS-2 folio and Ls numbers.

If we follow our previous example, and supposing the original home has added an additional member and the individual whose ls = 3 has left the home and created a new household, the individual pid_link is as follows for their respective homes:

At the original home:

Folio 2002	Ls 2002	Folio 2005	Ls 2005	Pid_link	Ls01a
00001000	01	00001000	01	0000100001	1
00001000	02	00001000	02	0000100002	1
00001000	03	00001000	03		3
00001000	04	00001000	04	0000100004	1
-	-	00001000	05	0000100005	4

At the new home:

Folio 2002	Ls 2002	Folio 2005	Ls 2005	Pid_link	Ls01a
00001000	03	00001003	01	0000100003	4

-	-	00001003	02	0000100302	4

As you can see, the pid_link for new individuals was added to the original and to new households. It is comprised of the MxFLS-2 folio and Ls numbers being that during the previous survey round, these individuals were not a part of the sample and as such, had no folio or Ls assigned to them. For new homes created by the individual who left the original home and who had an Ls = 3, at his new home he was given a pid_link made up of the original survey folio and Ls numbers.

In the c_ls section (Household members List) from Book C (Log) we can identify the individuals who are no longer home members with the ls01a variable. The pid_link was left blank at the original home in cases where individuals had left their original home to form a new one (as can be seen in the previous example), due to the fact that no new information on those individuals could be provided.

When the level of observation is different from the home or individual [observation], usually this is because the database was collected as a network consisting of a group of repeated questions pertaining to a series of events or specific themes. For example, in the health and family planning section (SP) in book C, each observation corresponds to a specific healthcare provider or clinic, and there are a vast number of observations per home. In this database, the folio and sequence combination identify each observation. The variable which identifies the themes or events in each database is called a sequence.

Occasionally, for individual questionnaires, there are databases where the information is repeated again by event or theme as is the case with the External Consultation section (CE) in book 3B. In order to have one single identifier for each of the observations in the second file (iiib_ce1), you would need to use the folio, Ls and sequence variables for this database.

5.3. Household Questionnaire Files and Code Manuals

- There are twelve questionnaire files, one for each household book (hh05q_byy); where letter q refers to the corresponding questionnaire files.
- There are twelve code manual files. One for each household book (hh05cb_byy); where the letters cb refer to the code manual files.
- The zipped file (hh05q_all) contains all twelve household questionnaire files.
- The zipped file (hh05cb_all) contains all twelve code manual files.

The questionnaire files and code manuals are in PDF format.

5.4. Community Files

Like the household surveys, the community archives include: database files, questionnaire files and code manual files. The community files are organized in the same way as the home survey files:

- There is a zipped file, (loc05dta_all) which contains all community survey files.
- In turn, two separate files contain information on each community survey book (loc05dta_byy). Each file corresponds to a community questionnaire.

Letters yy represent the specific book name: The Location characteristics Book (CC), the Book on Location Prices (p).

The questionnaire files and code manuals are in PDF format.

All files found in the database as well as other documentation, have been compressed in zipped files for an easier download. WinZip software is required in order to access each zipped file from your computer. WinZip software can be downloaded for free at: http://www.winzip.com/downwz.htm.

6. File Combinations

For better management, the MxFLS-2 data is stored in separate files (for example, section ii_ah in book 2 contains information about assets and section ii_in also found in book 2, contains information on non-employment earnings). To analyze the MxFLS-2 data, combinations between the different data files is usually required. The manner in which data is combined depends on the nature of the analysis taking place. This section discusses some of the ways in which data files may be combined.

6.1. Linking Data

Sometimes, you may need a group of observations found in different files while analyzing. For example, book 3A section iiia_ata and book 5 section y_atn, both contain data on time allocation. Data in section iiia_ata refers to adult time allocation and section y_atn refers to child time allocation. The adult variables begin with ata, while child variables begin with atn, both referring to the same information. This being the case, it might be interesting to combine these two variables for both age groups. To do so, it is necessary to use the APPEND command for STATA or the SET command for SAS. The end result will be a file containing all child and adult observations. It is worth mentioning that the variable names are different from files you may wish to link, some file variable could be re-named in order to match them with another file.

For the most part, different files may be linked. As a rule of thumb however, you should verify if any section was included in the Proxy books while using data from books 3A, 3B, and 4. This way, the interviewers' data (answered by him or her self) could be combined with collected data in the Proxy book for other individuals.

The following table shows additional combinations. You should keep in mind that some files may need to be restructured before they are combined due to the different levels of observation. Also, some files may require you to re-name some variables.

Table 4. Home Files that allow Links		
Theme	**Files**	**Specific Informant**
Files	ii_ah and iiia_ah	Book 2 and Book 3A Informant
Credit	ii_crh and iiib_cr	Book 2 and Book 3A Informant
Non-employment earnings	ii_in and iiia_iin	Book 2 and Book 3A Informant
Crime and Violence	ii_vlh and iiia_vli	Book 2 and Book 3A Informant
Educational History	iiia_ed and v_edn	Children and adults
Employment decisions	iiia_tb and v_emn	Children and adults
Time allocation	iiia_ata and v_atn	Children and adults
External Consultation	iiib_ce and v_cen	Children and adults
State of Health	iiib_es and v_esn v_he and v_esn	Children and adults
Hospitalization	iiib_hs and v_hsn	Children and adults
Self-Help	iiib_ats and v_autn	Children and adults
Cognitive Abilities	ea_eca and en_ecn	Children and adults

6.2. Combining files at the Individual, Household and Community Level

One of the most basic ways to create file combinations involves the counter crossing of a given person's files (on an individual level). If both files contain data on the same level of observation, the cross will be "one to one".

6.2.1. Combining files at the Individual Level

Suppose your objective is to analyze a file containing information on education, and an individual's height and weight measurements. Book 3A file iii_ed (adult) and Book 5 file y_edn (children) contain educational level information. Book S file s_sa contains weight and size observations for each individual. To create this new file, we organize each of the three files by order of folio and Ls and subsequently we merge by folio and Ls.

For example, suppose that a researcher wants to create a file with the child's current weight and the child's weight at the time of birth. Book S file s_sa contains current weight information, while the iv_he2 file contains weight information at the time of birth. There is one observation per individual in s_sa just as in iv_he2. However, observations in s_sa are identified using folio and Ls, and observations in iv_he2 are identified using folio and he06 (remember, book 4 section HE inquires on history of pregnancies and identifies each pregnancy with a personal identifier Ls for each son or daughter that is a household member).

First it is necessary to remove all adult information from s_sa (remember that this file contains information on both children and adults). This can be done by eliminating observations by those who are over the age limit that is desired. Then it would be necessary to change the he06 name to the Ls in iv_he2 (or the Ls to he06 in s_sa). Organize the files by folio and Ls and merge by folio and Ls. It is important to keep in mind that if the universal observations in s_sa are greater than the universal observations in iv_he2, there will be observations which will not coincide and as a result will have to be erased.

6.2.2. Crossing two files at the Household Level

For two files at the household level, such as ii_in and ii_inr from book 2, which contain household characteristic data, we organize each file by folio and merge by folio.

Crossing two files at the infrastructure level from community observations: For two files at the community level, such as esc_dg and esc_de (school characteristics) we organize each file by folio and merge by folio.

Crossing two files at the community level: Suppose that the researcher wishes to create a file with information on common school characteristics contained in esc_dg and the characteristics of small health care providers contained in peq_gp2 for each community. In this case, before merging these two files, it is necessary to do a merge by each folio file with the front cover of its corresponding book. This is important due to the addition of a new variable id_loc, which identifies each community and will be the variable used for the merge. Once you have this for each file, each file will be organized by id_loc and the merge will then be done by id_loc.

6.3. Combination of files at the Individual and Household Level

Often, it is necessary to counter-cross different files that are not on the same level of observation. For this type of situation, it is sometimes necessary to restructure the different files. When looking for the best way to combine distinct MxFLS-2 files it would be useful to determine if the variables identified in one of the files are a sub-component of variables identified in another file.

For example, suppose we need to create a file containing information on the level of education and adult household assets. The identifying variables in the education file (iiia_ed) are folio and Ls, but the identifying variable on household asset information (ii_ah) is only folio. In this case, given that the identifying variable in ii_ah is a sub-component of the identifying variable in iiia_ed, you could simply create a merge using the folio variable. This is possible if the identifying variable for individuals in iiia_ed is in individual level. This will result in a database at the individual level where the information on household assets is repeated for all the individuals who are members of the same household.

Now suppose that we need to merge (cross) the literacy file (book 3A) and the individual crime and victimization file in book 3A. The identification variables in iiia_ed are folio and Ls. The identifying variables in iiia_vli1 are folio, Ls and SEQUENCE. An individual could have many recorded assaults or violent incidences he or she has experienced in iiia_vli1. The file we want to create would be done in the following way:

First, given that the identifying variables in iiia_vli1 are sub-components of the identifying variables in file iiia_ed, you could simply combine by folio and Ls. This

would result in a database containing as many logs as violent incidences reported by each individual. Each log contains information on the literacy level of each individual and information on a particularly violent incident.

The second option is to restructure the iiia_vli1 file in such way that it is organized by individual level as opposed to the level of violent incidences reported. In this case, the identifying variables would be folio and Ls. This could involve the creation of a file containing, for example, variables vli(01) – vli(15) for 1 victimization incident, vli12_1_1, vli12_1_2, …, vli12_1_15 (in this case it would be 15 because the highest number of violent incidents reported by one individual is 15). This file will have more variables than iiia_vli1 but would have a lesser amount of observations. If the data in iiia_vli1 is restructured to create a file at the individual level, the combined files by folio and Ls with the iiia_ed file will be created at the individual level and will contain information on literacy and all information pertaining to victimization.

Finally, it is worth mentioning that the restructuring of data files in order to convert them to a different level other than their original one, could easily be done on STATA with the reshape command or in SAS with the proc transpose command.

6.4. Combination of Household data with Community data

When combining household survey data with community survey data, two basic cases arise: 1) household data at the community level and 2) household data at the level of infrastructure

Combination of household data with community level data: Suppose that a database is required with information on children's state of health (found in book 5 section v_esn) and the price of child medicine found in the Community Price book section pre_prf2. The first thing we must do in order to create this combination of files is create a combination of file v_esn with the front cover of Book C by folio. This is done in order to add the location identifier id_loc to the file (this variable is found on the front cover of Book C). Similarly, create a folio merge of files pre_prf2 with the front cover of the Community Price book (variable id_loc is found on the front cover of the Community Book and household Book C). Once the id_loc variable is found in both files, we can create a merge using the id_loc variable.

7. Using MxFLS-1 information with MxFLS-2

This section provides a guide for the MxFLS-1 and MxFLS-2 databases and assistance in obtaining a longitudinal database of homes and individuals.

7.1 Identifier differences in MxFLS-1 and MxFLS-2

In MxFLS-1 the folio, known as the household identifier, is comprised of 8 digits of which the last two digits are zeros.

x x x x x	00
FOLIO	Ending

In MxFLS-2 the folio is also comprised of 8 digits but unlike the folios in MxFLS-1, these may have a different ending other than zero. If the household interviewed is an original home, the folio is the same and the last two digits will be zero. However, if an MxFLS-1 original member has departed from the home and forms his/her own household, that individual's folio will be made up of the first six digits stemming

from the original home and the last two digits will be made up of the individual identifier in MxFLS-1.

x x x x x x	x x
FOLIO	Ls

For example, an MxFLS-1 home has the folio "00001000". During MxFLS-2, the family's son leaves the home to begin his own family. The individual who left the home had an individual identifier Ls "03" during MxFLS-1. The folio for the respective families would be as follows:

Folio	Type of Home	Observations
"00001000"	Original Home	Same home that was interviewed during MxFLS-1
"00001003"	New Home	Home created by individual with Ls "03" from previous household.

Given the number of original homes unfolding during MxFLS-2, when a new home was formed, the individual identifier (of the original individual) also changed and is not the same Ls he had at the original home.

Continuing with our previous example, the individual who formed his new home does not have the same Ls he had at the original household.

Type of Home	Folio 2005	Ls 2005
Original Home	"00001000"	"03"
New Home	"00001003"	"01"

In order to have an individual identifier in both databases, the identifier pid_link was created, allowing individual identification in both survey rounds. The pid_link is comprised of the individuals' original folio and ls identifiers, that is to say, the folio and Ls corresponding to MxFLS-1, regardless if the individual is in the original or new household during MxFLS-2. The pid_link is made up of 10 digits, the first 8 digits are the folio digits and the last two correspond to the LS.

x x x x x x x x	x x
Folio	Ls

For new individuals who are added to the original homes or who are added as a new home member, their pid_link will be made up of their existing home folio and respective Ls identifier.

For example, continuing with the same folio we have been using, suppose that during MxFLS-2 we discover the unfolding of a new home by a member who has left the original home to form his own (consisting of two individuals), and a new household member was found at the original home. Their pid_link for each one would be as follows:

Type of individual	Folio MxFLS-2	Ls MxFLS-2	Folio MxFLS-1	Ls MxFLS-1	Pid_link
Head of Household	"00001000"	"01"	"00001000"	"01"	"0000100001"
New Member Household	"00001000"	"04"	-	-	"0000100004"
Member who has left home	"00001003"	"03"	"00001000"	"03"	"0000100003"
New Member of New Home	"00001003"	"02"	-	-	"0000100302"

7.2 Combining MxFLS-1 information with MxFLS-2

Before combining information between MxFLS-1 and MxFLS-2, you would need to create databases with all the information required in each round according to previously mentioned explanations. Once the databases are ready, the information can be added at the individual, home or community level.

7.2.1 Combination of data at the Household level

There are various options available in combining household level information from both survey databases. If you only wish to combine the information from both databases to the original MxFLS-1 households, you could add the information by folio. This way, you would only be adding MxFLS-2 household information to the same, original homes interviewed during MxFLS-1.

If what you require is adding MxFLS-1 home of origin information to MxFLS-2 households, regardless of original or new home status, then you need to create a variable composed of the first 8 pid_link digits and then carry out the counter-cross. The pid_link is used (as apposed to the folio) because there are original individuals within the database who have formed a new home with an other individual (who could also be an original individual from our sample) and in forming this new home, one of their folio's will be used to create this new folio identifier. The way to track the individual's home of origin is through the pid_link.

In STATA, the way to create the original folio from the pid_link is as follows:

gen str8 *var* = substr (folio, 1,8)

In contrast to MxFLS-2, the folio from the first survey round is numeric and in order to carryout the adding and combination of both databases, it is necessary to have the same format for both folios. To turn the MxFLS-1 folio to STRING using STATA, you could use the following instruction:

gen str8 var1=string(folio, "%08.0f")

Where var1 is the newly created variable.

7.2.2 Data combination at the Individual Level

If the combination required is at the individual level, then this should be done using the pid_link. In MxFLS-1 databases the pid_link is not created, for which they must

be generated as explained below in order to subsequently create the merge with this variable.

STATA commands in order to create MxFLS-1 pid_links:

$$
\text{gen str8 } var1 = \text{string (folio, "\%08.0f")}
$$
$$
\text{gen str2 } var2 = \text{string (ls, "\%02.0f")}
$$
$$
\text{gen pid\_link} = var1 + var2
$$

The *var1* and *var2* variables are created in order to generate the folio and Ls string of 8 and 2 digits respectively. Being that the MxFLS-1 format is numerical and when the folio is composed of less than 8 digits, it is necessary to fill-in to the left with as many zeros as needed to create it with the necessary 8 and with the Ls. When the Ls is composed of one digit, a zero is added beforehand so that it is composed of the two digits required.

Folio MxFLS-1	Ls MxFLS-1	Var1	Var2	Pid_link
1000	1	"00001000"	"01"	"0000100001"

7.2.3 Combination of data at the Community Level

If a combination at the community level is required, the adding-on of information from either database could be carried out using the folio variable, which can be found in the community database.

8. Number of questions and Variable names

The majority of MxFLS-2 variable names correspond to the survey's question number. For example, variable names for section ED (Education) begin with ED and with the corresponding question number. For MxFLS-2 questionnaires, there was an attempt to match the number of questions and names to MxFLS-1. If a question was added or changed in MxFLS-2, generally an "a" or "b" was added to the question.

8.1. Missing values and Special codes

For numeric variables, a period (.) means a missing value. For text variables (string), a space also means a missing value. These missing values found in some variables could be a result of a no-response to a given question, and as a result of the appropriate "skips" or filters within the questionnaire. For example, question ah01 in book 3A requests that the interviewer verify if the informant has completed and answered book 2. If it is so, this person will not need to answer this section and can skip to the next one. If the answer is yes (ah01=1) the interviewer will ask the entire AH section. Other instances where missing values can exist are when the interviewee does not wish to, or cannot answer the question. In this case there are special codes used ending in 5, 7, 8 or 9 depending on the situation.

These special codes can take the following values, depending on the question and space assigned for the response:

5, 95, 995 or 9995 = Out of range, response cannot be recorded as there is not enough space in the questionnaire.
7, 97, 997 or 9997 = Interviewee did not wish to respond
8, 98, 998 or 9998 = Interviewee did not know the answer

9, 99, 999 or 9999 = The question does not apply

8.2. Type of answers

The majority of MxFLS-2 questions require a numeric value response. However, there are some cases of open response questions. There are a maximum number of digits permitted for questions that require a numeric response. When a maximum number is reached in a given question, the interviewer must use the corresponding special code.

Some questions allow multiple answers. For example: Within your family, who decides matters regarding your children's education? In this case, alphabetical responses are specified. A different variable was created for each option so that new variables are given a value of 1 if the interviewee responds affirmatively and no value when the response is negative.

For categorical variables, the questionnaire provides different options for each answer. The Coding manuals also provide the response layout. For categorical variables, frequency distribution is also provided. When multiple answers are allowed in these specific variables, the Coding manuals specify the number of interviewees who answered each question. If a variety of answer combinations were possible, the Coding manual would not give the response layout.

9. Special characteristics of MxFLS-2 household information

In this section, data with special MxFLS-2 characteristics is presented, which could affect your analysis.

9.1. Symmetric Information

Husbands and wives answer the same information within two sections of book IIIA, assets (AH) and household decisions (DH). In other words, husbands answer questions related to himself and to the rest of his family (including his wife). Similarly, the wife will answer the same questions; that is, she answers according to her and to the rest of the household members (including her husband). This information allows comparisons on what a couple thinks of themselves and of other household members.

Book IIIA section AH (assets) is only answered by household members who have a partner and who have already provided household asset information requested in book II. If these two individuals are married and living in the same household, the asset information provided may be the same, and one person's information could be used to fill-in possible omission responses from the other. Similarly, in book IIIA section DH, married individuals describe the way household decisions are made. Information from both informants may be used to verify the consistencies and in case of missing information, it could be complimented.

The same thing occurs in book IIIB section TH where detailed information on non-resident household siblings is collected. We gather information on non-resident sibling details and characteristics for all household members.

9.2 Duplicate Information

There are some questions that are asked in multiple sections within the questionnaire. For the most part, the informant was a source of information and a

Proxy informant as well (or providing information on pre-printed lists). For example, book 3 in the household asset section (AH), contains information also requested on individual books. Although it may be easier to use the added information from the household book, it is more likely that the information provided in the individual books is the most veracious, or could be used to compliment information from the first source.

Age. Information regarding age is asked in book C section LS (for the most part one household members provides the age of all home members) and on the cover of each individual book. We do not correct possible age inconsistencies between what is reported in book C and that reported on the cover of individual books.

Sex. Information on an individual's gender is asked in book C section LS, book S section SA and on the cover of books 3A and 3B. As in the case of age, we do not correct any possible inconsistencies between what is reported in book C and what is registered on the cover of the individual books.

Marital Status. Marital status is asked in book C section LS and on the cover of books 3A, 3B and IV. In data cleaning, we try to be sure that the marital status recorded on the LS list was the same as the marital status logged on the front cover of the individual books

Level of Education. Book C section LS reports the highest level of education attained. This question is repeated for some individuals in book 3A and 5.

Employment and Non-employment earnings. Book 3A section TB requests information on employment earnings. This is also requested in the Proxy book. Book 2 section IN asks for information of non-employment earnings at the household level and book 3A section IIN at the individual level. Information at the individual level is preferred, but the added data at the household level is especially useful in analyzing the total household income in cases when there is missing household member information.

Status of Parent survival. The household member list is recorded in the LS section, which contains a line number for each household member's mother and father (ls06 and ls07). Special codes are used to record whether the parents live but are not household members or if they are deceased. Book 3 section TP (parents) directly inquires if the mother or father are living or deceased.

9.3. Relationships between Household members

MxFLS-2 contains information on family relationships, particularly between husband and wife and between parent and child. This is done in such a way that it becomes clear to identify who is whose partner and who is whose child. The information is not limited to household members, and also stems into non-resident relationships.

Kinships within the LS roster (Book C)

Section LS has complete information on the kinship of current household members. The information provided by the list of household members LS is as follows:

Table 5. Kinship Relations		
Variable	**Information**	**Observations**
Ls05	Who is the head of household in MxFLS-2 and what is the kinship associated with this member in relation to the other household members?	This information gives the kinship relationship each household member has with the head of household. For example, if members LS 3 and 4 are sons of the head of household, they would be brothers or half-brothers. If member 5 is the head of household's mother and member 3 is son of the head of household, it is almost certain that member 5 is grandmother to member 3. Other examples are not so definitive. For example, if members 6 and 7 are grandchildren to the head of household, they could be siblings or cousins.
Ls06 Ls07	Household member's Father line number (LS) Household member's Mother line number (LS)	This variable allows you to link an individual with his or her parents through their individual id. For example, if a person X has an ls06=1 and an ls07=2, her father is the head of household (given that the head of household is always LS=1) and her mother is LS=2.
Ls08	Number line pertaining to the caregiver responsible for children under the age of 15.	This variable indicates a caregiver within the household for children under the age of 15

Identification of children in other sections

Book IV sections HE1 and HE2 contain information on the history of pregnancies for women between the ages of 14 and 49 years of age. The data collected includes specific details about the child since the moment of birth and the mother's pre and post-natal care.

These sections correlate each birth with the personal identification number (LS) for each son or daughter who are household members. This allows a counter-cross of information on children born [elsewhere] with information on the children (sons or daughters) found in the survey. Other son or daughters whose LS is unknown where coded with a "99".

Sons or daughters who are non-household members Book 3B section THI provides detailed information on non-household-member sons or daughters. Detailed household child member information was not collected if another informant has provided information. For example, if person X (a man) is married to Y (a woman) who is father to the same children as Y and has had no other wives, in this case, the information provided by person Y regarding their children is used as information for children belonging to person X.

If person X had children with another woman, person X would answer questions regarding his other children.

Non-household member siblings Book 3B section TH (siblings) gives detailed information on siblings who are not household members. In this case, and contrary to what happens with non-household member children, non-household member

sibling information of all household members was collected (even though non-household member sibling information could have been provided by another member).

It is necessary to counter cross different sections in order to calculate the total number of children, siblings, information on the informant's parents or information pertaining these kinships characteristics. The following table shows you how to create this cross of information.

Table. 6 Source of information for an Informant's immediate family members	
Siblings	
Within the home	• For a household member, use variable ls05 in book C section LS to identify household members who are siblings. For any other member whose mother or father is a household member, check the Household Member roster (LS) for other members who identify their parents in ls06 and ls07 in book C section LS. For a household member who is not head of household and whose parents are not household members, sibling information is nuclear. • Once household siblings have been identified, you can consult the information on individual books.
Outside the home	Use the sibling characteristic information pertaining to non-household members in book IIIB section THI.
Sons or Daughters	
Female informants	For each childbirth, use the information in book 4, section HE. For living children, you can also use information from book 5.
Male informants	For each child belonging to the current couple, use information given by the wife/spouse from book IV section HE and book 5. For children who are non-household members over the age of 15 from a previous spouse, use data from book 3B section THI (children).
Parents	
Within the home	Depending on the required information, use data from books 3A and 3B from each parent or from

	book C section LS.
Outside the home	Use the information from book 3B section TP (parents).

10. Cleaning MxFLS-2 data

This section describes the systemization process of MxFLS-2 information.
Collecting data, systemization and cleaning data collected by a team of researchers are all fundamental processes in a given survey. In the case of MxFLS, the process of entering data was carried out almost immediately after collecting it in the field.
Once the questionnaire design came to conclusion, CAFÉ (Computer Assisted Field Editing) was designed as a system for entering data. CAFÉ was installed in laptops and in our editor's equipment, who would travel to the designated location with our researchers and who would supervise the collection of data and information.

This way, once our interviewers filled out the household information on the pre-printed questionnaires and once a fieldwork supervisor had reviewed it, this was handed over to the Data Entering team. The process of entering data while still working in the field would permit, first, the possibility of returning to the home to finish and/or verify any errors, missing information or inconsistencies and secondly, make the cleaning process a more efficient and rapid experience.

The editing team incorporated and corrected the information directly form their laptops using the assisting editing information system designed to detect any errors or omissions. For example, if an informant's gender recorded in the LS roster does not coincide with the gender recorder on the cover of book 3A, a message of error is generated. The editor would be held responsible for discussing possible errors with the interviewer. With our data entering system on the field, some errors could easily be corrected. In our previous example, the interviewer could easily recall the interviewee's gender as the interview most likely took place a few days ago. Other errors require that the interviewer return to the home and verify possible errors or omissions directly with the informant.

CIDE and UIA carried out an additional data cleaning process to correct any final errors and to make the data public domain.

10.1 Search System and cleaning process: "Look Ups"

To detect and resolve more complicate errors, a cleaning process called "Look Ups" (LU) was implemented. Sophisticated Computer programs were required to guarantee the quality of information while using LU. This process made use of additional information contained in the questionnaires that was not entered into electronic format. For example, and inconsistency could have been created by an inadequate correction; this was corrected by a side note reference created by the interviewer, this way data is verified and corrected.

As an example, the LU program verifies that parents are at least 12 years older than their children. The program will automatically generate an error detection that can be corrected in a transparent way and verified using the original questionnaires. However, if the specialist is not certain on how to correct the data, then the information is not corrected immediately. Some problems were relatively

easy to correct, others such as filters or omitted skips could not be corrected because that specific data had not been collected.

10.2 Special cleaning for Open questions

Open questions. Some questionnaire questions allow open questions. A specialized trained team carried out data cleaning for these open answers.

Variable responses of "other, specify". This type of answer was found when it did not correspond to the options available on previously coded answers. In cleaning answers of "other, specify", it was necessary to verify the text answers and decide if another answer could be added to an existing category, or if an new category was needed. Specialized trained teams from CIDE and UIA carried out this cleaning. New categories were created when some answers were substantially different to pre-coded answers.

Therefore, users have the option of adding data. A more refined adding and removing of data would be impossible if the new codes would not have been created. These modifications will be considered when elaborating survey tools (questionnaires) for future MxFLS rounds.

Three types of "other, specify" answers where cleaned:

- Simple questions allowing only one answer: For this type of questions, the "other" answers were re-coded as a new or existing answer category. For example, question 11 in book C section CV inquires if the household disinfects their drinking water. The original questionnaire allowed 5 categories and the "other" option. After the cleaning process, a new category was added: "purified water purchased".

- Questions where multiple answers are permitted: The "other" option was re-coded as a new or existing category and eliminated. For example, question GH02 in book 3B section GH inquires on what type of beverages are preferred while eating. The original questionnaire allows 4 categories (water, beer, soda and tequila/*pulque* [fermented Maguey drink]) and the "other" option. After the cleaning process, two new categories where added (rum and hot drinks) and the "other" option was divided into "other alcoholic drinks" and "other non-alcoholic drinks".

- Questions related to matrix. The cleaning of "other" answers could have created another specific category (another option) within this section. For example, book 3b section CA asks who provided health insurance for the individuals. In this section was created other option "Popular Health Insurance".

At the end of the data cleaning process, some observations that withheld the "other" option remained in a text format category. These variables are not available to the public.

Numeric Variables. Some numeric questions did not fit in the space provided. In such cases, our trained interviewers were asked to fill-in as much space as provided and end with a code 9 ending in 5 ("lack of space") and record the correct response in the questionnaire section "observations" or in "other, specify". If the interviewer provided a valid justification for this answer, numeric fields were added, allowing the correct response to be entered and "lack of space" was replaced with the correct answer. It was not possible to correct all answers with this code, so special codes will sometimes appear within the data.

10.3. Verification Section

In all sections, there was an attempt to:

- Verify the LU process and determine if any remaining errors or inconsistencies could be corrected.

- Verify numeric answers for the purpose of special codes and verify all variables with empty answers or "don't know" codes.

- Create or correct the "_1", "_2" variables in such a way that special codes could be preserved and the number associated with a particular variable could only contain valid answers. Variables "_1" are generally associated with a numeric value and are indicative of whether or not the person could answer the question.

- Verify that "skips" or "filters" were followed accordingly, and make appropriate corrections in cases where missing data was not due to skips or filters.

- Designate names and variable labels as clearly as possible.

- Correct the possibility of duplicate observations.

- Locate and eliminate any variable that could identify the informant.

10.4. Verification of identifiers through books

It is essential that the identification numbers FOLIO and LS be assigned correctly. As such, the assignation of identification was rigorously verified using a special program, which used the same identifiers throughout all sections. This program verified that the age and name of each individual were correct throughout the different sections.

10.5. Verification on units of measure

Some questions contained a numeric response allowing a unit of measure. For example, question CS10 from book I (Household consumption of meals) asked about the amount of products bought within the last few days, having as a response the options of kilograms, pieces, or other units. The informants could have clearly provided wrongful answers. For example, if an individual answered he had bought 10 kilograms of bread; the appropriate unit of measure for this response is more likely to be 10 pieces instead of kilos. In this same way, if a woman reported an abortion after 11 months of pregnancy, we accordingly judge that the appropriate unit of measure intended was weeks as opposed to months. Such corrections were applied to very few cases.

10.6. Creating variables and files

Some variables and data files were created for the purpose of an easier use, for example:

Answers associated with occupation and employment sector in book 3A were coded but also allowed open answers. These open answers were subsequently coded to make them comparable with CMO codes for the occupation of activity

classifications. These new variables are '… cmo' and are followed by the corresponding number question.

Information regarding age, date for birth as well as educational information could vary between books. We created our "best guess" of these variables for people using the MxFLS-1 data, which can be found in the BESTGUESS? database. For book 4 (reproductive health) we substituted the names of children provided by their mothers with the corresponding identification number belonging to each child. This could only be done for children who lived with their mothers and whose mothers were between the ages of 14 and 49 years of age. This is the case of variable he06 contained in section HE1 and of variable he19a found in section HE2. These variables were assigned a "99" code for children who were not at home.