Instructions to Generate and Download Datasets to Accompany the   
IPUMS PMA Longitudinal Analysis Guides   
for R and Stata Users

December 2022

This page is intentionally blank.

# Table of Contents

[Preface 1](#_Toc122094472)

[Introduction 3](#_Toc122094473)

[1 Generate the pma\_00001 Dataset 5](#_Toc122094474)

[2 Generate the pma\_00002 Dataset 24](#_Toc122094475)

[3 Generate the pma\_00003 Dataset 26](#_Toc122094476)

[4 Generate the pma\_00004 Dataset 30](#_Toc122094477)

[5 Generate the pma\_00005 Dataset 32](#_Toc122094478)

[6 Generate the pma\_00006 Dataset 34](#_Toc122094479)

[7 Generate the pma\_00007 Dataset 36](#_Toc122094480)

This page is intentionally blank.

# Preface

This document provides instructions for extracting datasets from the [IPUMS PMA website](https://pma.ipums.org/pma/)[[1]](#footnote-1) so you can practice the analyses described in the *IPUMS PMA Longitudinal Analysis Guides* – either for R or Stata users. Those guides and the programs that accompany them are freely available to download from a [GitHub repository website](https://github.com/IPUMS-Global-Health/IPUMS-PMA-Longitudinal-Guide)[[2]](#footnote-2). Although there are two guides, one for R and one for Stata, the material in one is largely a mirror of the other, and so hereafter in this document we refer to “the guide” meaning generically both guides.

The guide uses seven different datasets in its examples – each with different characteristics. We are not able to provide the datasets for direct download because of PMA permissions policies, so as a next best solution, this document shares step-by-step instructions so you may generate whichever datasets interest you.

**Suggested citation**

Mia Yu, Matt Gunther, Dale Rhoda, and Devon Kristiansen. Instructions to Generate and Download Datasets to Accompany the IPUMS PMA Longitudinal Analysis Guides For R and Stata Users (December 2022). Minneapolis, MN: IPUMS. pma.ipums.org

**Abbreviations**

GZ Extension for a file compressed using the [Gnu Zip](https://www.gnu.org/software/gzip/) archiving algorithm

IPUMS [Integrated Public Use Microdata Series](https://www.ipums.org/)

PMA [Performance Monitoring for Action](https://www.pmadata.org/)

**Revision History**

Revisions to this document are listed by date and accompanied by comments [here](https://github.com/IPUMS-Global-Health/IPUMS-PMA-Longitudinal-Guide/commits/main). Questions and suggested changes are welcome! Please submit requests to our [Issues](https://github.com/IPUMS-Global-Health/IPUMS-PMA-Longitudinal-Guide/issues) forum on GitHub.

This page is intentionally blank.

# Introduction

To follow these instructions, you need to have an IPUMS PMA account. To obtain one for free; simply click the link that says ‘Register’ at the top of the page at <https://pma.ipums.org/pma/>.

This document shows thorough step-by-step detail to generate the dataset named pma\_00001 in the next chapter. The subsequent chapters describe the variations on that same process for requesting datasets 00002 through 00007.

Throughout this guideline, there are three different kinds of marks on the screenshot images.

The red circle indicates where you need to click to accomplish a step.

The light orange circle indicates an alternative way to accomplish a step.

The green square points out where you need to check but no actions are   
 necessarily needed.

Finally, instructions with yellow hightlights differ in some way from those in earlier chapters, so pay special attention to what they say.

This page is intentionally blank.

# Generate the pma\_00001 Dataset

1. Log in to IPUMS PMA and navigate to the the front page of thewebsite.   
   Click “Get Data” button in the “CREATE AN EXTRACT” section.

(Clicking “SELECT DATA” near the upper right corner will work too.)

Graphical user interface, website

Description automatically generated

1. Select the topic for data browsing. For this dataset, select “FAMILY PLANNING” under the “PERSON” unit.

Graphical user interface

Description automatically generated

(If you have selected data before, this page might be skipped. This is fine as you can change the topic on the next page as well.)

1. The next page should say “CURRENTLY BROWSING: “FAMILY PLANNING – PERSON” near the upper left corner. If not, click “CHANGE” to change the topic.   
   Click the purple “SELECT SAMPLES” button:

Graphical user interface, application

Description automatically generated

1. IPUMS provides two kinds of samples: “Cross-sectional” and “Longitudinal”.   
   For pma\_00001, you need to select "Longitudinal” and “Wide” format:

Graphical user interface, text

Description automatically generated

1. Select the samples you want.   
   For pma\_0001, you can simply select “All Samples (wide)” and then click “SUBMIT SAMPLE SELECTIONS”.

Graphical user interface

Description automatically generated

1. The next step is to select variables wanted. One easy way to do this is to select variables via names. Click the “A-Z” dropdown bar under the “SELECT VARIABLES” section. You can also use the search function to search for the variable names that we list in step 6.

Graphical user interface, application, Teams

Description automatically generated

Here I’m going to use “EAID” as an example so I will select “E”.

Graphical user interface

Description automatically generated

Now you will see a page with variables starting with “e”. To add any variable you want to the list that will be included with your dataset (which is sometimes called an *extract*), simply click the “+” button that appears in the left-most column, before the variable name.

Graphical user interface

Description automatically generated

To have the exact same pma\_00001 dataset that we used to develop the guide, add the following variables from this alphabetized list to your cart.

AGE, AGEHQ

CONSENTFQ, CONSENTHQ, COUNTRY, CP, CROSS\_SECTION

EAID, EAWEIGHT, EDUCATT, EDUCATTGEN, ELIGIBLE

FQINSTID, FQWEIGHT

HQWEIGHT

INTFQMON, INTFQYEAR

MARSTAT

PANELWEIGHT, PANELWOMAN, PREGNANT

RESIDENT, RESULTFQ, RESULTHQ, RESULTLFQ

SAMEDWELLING, SAMPLE, STRATA, SUBNATIONAL, SURVEYWILLING

URBAN

WEALTHQ, WEALTHT

YEAR

Note that only the variables whose names are underlined are used in the code examples in the longitudinal analysis guide.

1. After adding the last variable, (YEAR for this dataset), click the “VIEW CART” button at the upper right corner.

Graphical user interface

Description automatically generated

This will direct you to the “data cart” that should look like this:

Graphical user interface

Description automatically generated

The cart lists all the variables you added and includes some additional information like “variable label”. You are free to add more variables and samples or unselect variables. Once you are ready to “check out”, click the “CREATE DATA EXTRACT” button.

1. Here is where the instructions differ slightly depending on whether you want to download the file to use with Stata or with R. You can also take an extra moment and download the extract in both formats. On the “EXTRACT REQUEST” page, the DATA FORMAT will default to whatever format you downloaded most recently. Click the “Change” link to set it to something else.

Graphical user interface, application, website

Description automatically generated

The screenshot below shows the user selecting “Stata (.dta).   
If you are working in R, select “Fixed-width text (.dat)”.   
You may only select one option at a time. Click “APPLY SELECTIONS”:

Graphical user interface, application, Teams

Description automatically generated

1. If you follow this process numerous times, your account will list each extract you have requested. Take a moment to enter a brief description of what you were accomplishing here. Maybe say “Recreate ipums\_0001 dataset”.   
   When ready, click the “SUBMIT EXTRACT” button at the bottom.   
   (Note in the green box below that if you want to use the exact same dataset used in the guide, this page should say that you have 6 samples and 34 variables in your cart.)

Graphical user interface, application, website

Description automatically generated

1. You will see a page like this that says “Processing…” under the “Formatted Data” column.

Graphical user interface, application

Description automatically generated

You will receive an email when your extract is ready – the process usually takes a very few minutes.

1. Once you receive that email, refresh this page and the formatted data status will change from “Processing” to “Download Stata”. Click the green square to download your data!!

Graphical user interface, application

Description automatically generated

The dataset will be formatted as a GZ file and you will need to use software to “unzip” the folder and extract the data file. As mentioned on the [IPUMS EXTRACT INSTRUCTIONS webpage](https://pma.ipums.org/pma/extract_instructions.shtml):  
*This file needs to be decompressed. Mac OS X will decompress the file when you double-click it. Most Windows users will have some type of application that unzips compressed files, but a significant number may not.* *If your Windows-based computer does not know how to decompress the file, you need to download decompression software. A free option is* [*7-Zip*](http://www.7-zip.org/download.html)*; other programs are also available.*

Important note !!!

The IPUMS website will name your datasets, starting with pma\_0001 for your first extract, 0002 for the second, etc. It’s quite possible that when you do this work to recreate what our longitudinal analysis manual calls pma\_0001, yours will have a different number in its filename. Don’t panic. You may copy the dataset that you unzip to a folder where you will do your longitudinal work, and then rename it to match the filename we used in the guide (and in the source code that accompanies the guide via the GitHub repository).

1. If you download the data and then learn that you selected the wrong file format setting, no worries! You don’t need to reselect everything. You can click the “revise” option under “Revise Extract”. You can also use this option to download the same data in a second file format if you are following along with the longitudinal guide steps using both Stata and R.

Graphical user interface, application

Description automatically generated

Clicking “revise” will bring you back to the page before you submitted the extract. From here you can change the sample, the variables, the format, the structure, and the sample members. IPUMS will auto-generate a note that describes this new extract. You can edit this note to what you want as well. Once you finished making changes to the dataset, click “SUBMIT EXTRACT” just like before. In the screenshot below, we are revising our first extract, as an example:

Graphical user interface, application, website

Description automatically generated

This page is intentionally blank.

# Generate the pma\_00002 Dataset

To create the other six datasets that are used in the longitudinal guide, repeat the multi-step process described in chapter 1. In this chapter and those that follow, we highlight only the selections that differ from those you made when generating pma\_0001.

For pma\_0002, the only instruction that differs, is step 6. Follow the other steps as described above for pma\_0001.

1. To have the exact same dataset that we used to construct the pma\_0002 examples in the guide, select the variables listed below.

ABORFOLLOWUP, AGEHQ, AVAILABLEFQ, AVAILABLEHQ

CONSENTFQ, CONSENTHQ, COUNTRY, COUNTRYSTR, CP, CROSS\_SECTION

EAID, EAWEIGHT, ELIGIBLE, ENUMID

FPCURRUSE, FQACQUAINTED, FQINSTID, FQWEIGHT

GEOBF, GEOCD, GEOKE, GEONG

HHID, HHIDORIG, HHIDORIGP1, HHMEMPRESP1, HHIDORIGP1DWELL, HHMEMSTAT, HHMEMUPDATE, HHNEWYR, HHNUM, HHOCCUPIED, HHPANELP1, HHPANELP2, HHREPLACED, HHSTATUS, HHTYPE, HHVACANTWHY, HQWEIGHT

INTFQMON, INTFQYEAR, INTHQMON, INTHQYEAR

LINENO, LINEORIG

MARSTATHQ, MCP

NEWPANELFEM

PANELWEIGHT, PANELWOMAN, PERSONIDORIG, PHONEOWN1, PHONEOWN2, PHONEWILLING2, PHONEWILLINGHQ, PREGNANT

RELATE, RESIDENT, RESPONDENT, RESULTFQ, RESULTHQ, RESULTLFQ, RESULTLHQ, RESULTPANEL

SAMEDWELLING, SAMPLE, SEX, STRATA, STRUCTURNO, SUBNATIONAL, SURVEYWILLING

TCP

UNMETNEED, URBAN

VISITNUMFQ, VISITNUMHQ

YEAR

Only the variables that are underlined are used in the code in the guide.

This page is intentionally blank.

# Generate the pma\_00003 Dataset

For pma\_00003, steps 4-6 differ from the instructions for pma\_0001.

1. IPUMS provides two kinds of samples: “Cross-sectional” and “Longitudinal”.   
   For pma\_00003, you need to select " Longitudinal” and “Long” format.

Graphical user interface, text

Description automatically generated

1. Select the samples you want.   
   For pma\_0003, you can simply select “All Samples (long)” and then click “SUBMIT SAMPLE SELECTIONS”.

Graphical user interface, text

Description automatically generated

1. To have the exact same dataset we used to write the guide, select the variables from this alphabetized list:

AGE

CONSENTFQ, CONSENTHQ, COUNTRY, COVIDCONCERN

EAID, ELIGIBLE

FQINSTID, FQWEIGHT

GEOCD, GEONG

PANELWOMAN, PHASE, PREGNANT

RESIDENT, RESULTFQ, RESULTHQ

SAMPLE, STRATA

YEAR

Only the variables that are underlined are used in the code in the guide.

This page is intentionally blank.

# Generate the pma\_00004 Dataset

For pma\_00004, only step 6 differs from the instructions for pma\_0001.

1. To have the exact same dataset used in the guide, select all the variables from the alphabetized list below.

AGE

CONSENTFQ, CONSENTHQ, COUNTRY, COVIDCONCERN

EAID, ELIGIBLE

FQINSTID, FQWEIGHT

GEOCD, GEONG

PANELWOMAN, PHASE, PREGNANT

RESIDENT, RESULTFQ, RESULTHQ

SAMPLE, STRATA

YEAR

Only the variables that are underlined are used in the code in the guide.

This page is intentionally blank.

# Generate the pma\_00005 Dataset

For pma\_00005, only steps 5 & 6 differ from the instructions for pma\_0001.

1. Next, you will need to select the samples you want. Here you want to click the “All Samples (wide)” option and change the sample members from “Female Respondents” to “All Cases”. Then click “SUBMIT SAMPLE SELECTIONS”.

Graphical user interface, text

Description automatically generated

1. To have the exact same dataset used in the guide, select all the variables from the alphabetized list below.

AGE, AGEHQ

CONSENTFQ, CONSENTHQ, COUNTRY

EAID, ELIGIBLE

FQINSTID, FQWEIGHT

GEOCD, GEONG

HHID, HHMEMSTAT, HHPANELP2, HHTYPE

PANELWOMAN

RELATE, RESULTFQ, RESULTHQ, RESULTLFQ

SAMEDWELLING, SAMPLE, SEX, STRATA, SURVEYWILLING

YEAR

Only the variables that are underlined are used in the code in the guide.

# Generate the pma\_00006 Dataset

For pma\_00006, only step 6 differs from the instructions for pma\_0001.

1. To have the exact same dataset used in the guide, the variables you want to add to the cart for this dataset are the following (all the variables have been sorted alphabetically):

AGE

BIRTHEVENT

CONSENTFQ, CONSENTHQ, COUNTRY, CP

EAID, EDUCATTGEN, ELIGIBLE

FPCURREFFMETHRC, FPPARTSUPPORT, FPPLANVAL, FPPLANWHEN, FQINSTID, FQWEIGHT

GEOCD, GEONG

MARSTAT

PANELWEIGHT, PREGNANT

RESIDENT, RESULTFQ

SAMPLE, STRATA

UNMETYN

YEAR

Only the variables that are underlined are used in the code in the guide.

This page is intentionally blank.

# Generate the pma\_00007 Dataset

For pma\_00007, only step 6 differs from the instructions for pma\_0001.

1. To have the exact same dataset we used to develop the guide, select the variables in the alphabetized list below.

CALENDARBF, CALENDARBFWHY, CALENDARCD, CALENDARCDWHY, CALENDARKE, CALENDARKEWHY, CALENDARNG, CALENDARNGWHY, CONSENTFQ, CONSENTHQ, COUNTRY

EAID, ELIGIBLE

FPBEGINUSEMO, FPBEGINUSEYR, FPCURREFFMETHRC, FPPLANVAL, FPPLANWHEN, FQINSTID, FQWEIGHT

INTFQMON, INTFQYEAR

KID1STBIRTHMO, KID1STBIRTHYR

LASTBIRTHMO, LASTBIRTHYR

OTHERBIRTHMO, OTHERBIRTHYR

PANELBIRTHMO, PANELBIRTHYR, PANELPREGENDMO, PANELPREGENDY, PREGENDMO, PREGENDYR, PREGNANT

RESIDENT, RESULTFQ

SAMPLE, STRATA

UNMETYN

YEAR

Only the underlined variables are used by the code in the guide.

1. <https://pma.ipums.org/pma/> [↑](#footnote-ref-1)
2. <https://github.com/IPUMS-Global-Health/IPUMS-PMA-Longitudinal-Guide> [↑](#footnote-ref-2)