





# CONTENTS

1 Longitudinal Data Extracts .....	1
1.1 Sample Selection .....	2
1.2 Variable Selection .....	6
1.2.1 Codes .....	8
1.2.2 Variable Description .....	11
1.2.3 Comparability Notes .....	12
1.2.4 Sample Universe .....	13
1.2.5 Availability Across Samples .....	14
1.2.6 Questionnaire Text .....	15
1.2.7 Checkout .....	16
1.3 Data for R Users .....	17
1.3.1 Select a Fixed-width File .....	18
1.3.2 Download .....	19
1.4 Long Data Structure .....	20
1.5 Wide Data Structure .....	23
1.6 Which format is best for me? .....	26



# 1 LONGITUDINAL DATA EXTRACTS

This chapter provides a guided tour of the [IPUMS PMA data extract system](#). While you may also access the original data directly from our partners at [PMA](#), harmonized data from IPUMS have a few additional features. For instance, you can request an extract that:

- includes samples from multiple countries
- includes samples from multiple rounds of data collection
- are formatted in either **long** or **wide** format

IPUMS PMA also makes it easy to switch between multiple [units of analysis](#) covered in PMA surveys. In addition to the data featured in this guide, you'll find surveys representing:

- [Service Delivery Points \(SDPs\)](#)
- [Client Exit Interviews conducted at SDPs](#)
- Participants in special surveys covering topics like [COVID-19](#), [nutrition](#), and maternal & newborn health

To get started with a longitudinal data extract, you'll need to select the **Family Planning** topic under the **Person** unit of analysis.

[Register here](#) to access IPUMS PMA data at no cost. See our [user guide](#) for details.

A video tour of the longitudinal extract system is available [here](#) on the IPUMS PMA Youtube channel.

The screenshot shows a web browser window for the IPUMS PMA website. The URL in the address bar is <https://pma.ipums.org/pma-action/variables/group>. The page title is "IPUMS PMA: vars by group". The main content area is titled "CHOOSE THE TOPIC FOR DATA BROWSING". It displays three categories: PERSON, SERVICE DELIVERY POINT, and INFANT. Under each category, there are two buttons: FAMILY PLANNING and NUTRITION. The "DESCRIPTION" link is located below each button. The "FAMILY PLANNING" button under the PERSON category is circled in red. The right side of the screen shows a "DATA CART" with 0 variables and 0 samples, and a "REVIEWS" section. The bottom of the page shows the full URL again: [https://pma.ipums.org/pma-action/variables/group?unit\\_of\\_analysis=person](https://pma.ipums.org/pma-action/variables/group?unit_of_analysis=person).

## 1.1 SAMPLE SELECTION

Once you've selected the **Family Planning** option, you'll next need to choose between cross-sectional or longitudinal samples. Cross-sectional samples are selected by default; these are nationally or sub-nationally representative samples collected each year dating backward as far as 2013.

The screenshot shows the IPUMS PMA website at [pma.ipums.org/pma-action/samples](https://pma.ipums.org/pma-action/samples). The top navigation bar includes links for LOG IN, REGISTER, GLOBAL HEALTH, and IPUMS.ORG. The main header features the IPUMS PMA logo and the text 'PERFORMANCE MONITORING FOR ACTION'. Below the header are links for HOME, SELECT DATA, MY DATA, and SUPPORT. The main content area is titled 'SELECT SAMPLES'. A note states: 'Variable documentation on the web site can be filtered to display only material corresponding to chosen datasets ([more information](#) on this feature)'. It says: 'You may select any of the below datasets for browsing. Please [log in](#) to see which samples you are authorized to include in extracts.' A radio button group for dataset type has 'Cross-sectional' (selected) and 'Longitudinal' options. To the right is a 'SUBMIT SAMPLE SELECTIONS' button. The main section is titled 'FAMILY PLANNING - PERSON'. It contains a 'All Samples' checkbox and a grid of checkboxes for years from 2015 to 2021. Below the grid, country-specific checkboxes are listed: 'Burkina Faso' (with 2021 P2, 2020 P1, 2018 R6, 2017 R5, 2016b R4, 2015 R2, 2016a R3), 'Congo (Democratic Republic)' (with 2020 P2, 2019b P1, 2018b R7, 2017b R6, 2016b R5, 2015c R4).

Longitudinal samples are only available from 2019 onward, and they include all of the available phases for each sampled country (sub-nationally representative samples for DRC and Nigeria are listed separately). You'll only find longitudinal samples for countries where Phase 2 data has been made available; Phase 1 data for Cote d'Ivoire, India, and Uganda can currently be found under the Cross-sectional sample menu (Phase 2 data will be released soon!).

Annual cross-sectional samples are also available for each of the countries participating in the new PMA panel study. See our [last post](#) for details.

Clicking the Longitudinal button reveals options for either **long** or **wide** format. You'll find the same samples available in either case.

**Important:** if you decide to change formats after selecting variables, your Data Cart will be emptied and you'll need to begin again from scratch.

The screenshot shows the 'SELECT SAMPLES' page of the IPUMS PMA website. At the top, there are navigation links for 'LOG IN | REGISTER | GLOBAL HEALTH | IPUMS.ORG'. Below the header, the IPUMS PMA logo is displayed, followed by 'PERFORMANCE MONITORING FOR ACTION', 'HOME | SELECT DATA | MY DATA | SUPPORT', and a search bar labeled 'Guest'.

The main section is titled 'SELECT SAMPLES'. It contains a note about variable documentation and dataset filtering. A message indicates that users can select any datasets for browsing and log in to see authorized samples. Below this, there are three radio button options: 'Cross-sectional' (unchecked), 'Longitudinal' (checked), and 'Wide' (unchecked). A red oval highlights the 'Longitudinal' option.

The 'FAMILY PLANNING - PERSON' section includes a 'Documentation' table with two columns. The first column lists sample types: 'All Samples (wide)', 'Burkina Faso', 'Congo (Democratic Republic)', 'Kenya', and 'Nigeria'. The second column lists corresponding years: '2020 - 2021', '2019b - 2020b', '2019a - 2020a', '2019 - 2020', '2019b - 2020b', and '2019a - 2020a'. An arrow points from the 'Longitudinal' radio button to the 'All Samples (wide)' row in the table.

The 'Sample Members' section contains four radio button options: 'Female Respondents' (checked), 'Female Respondents and Household Members', 'Female Respondents and Female Non-respondents', and 'All Cases (Respondents and Non-respondents to Household and Female Questionnaires)'. Below these sections are two 'SUBMIT SAMPLE SELECTIONS' buttons.

At the bottom of the page, there is a footer note: 'SUPPORTED BY: THE BILL & MELINDA GATES FOUNDATION, PMA, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA.' and a copyright notice: 'COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA.'

After you've selected one of the available longitudinal formats, choose one or more samples listed below. There are also several Sample Members options listed.

The screenshot shows a web browser window titled "IPUMS PMA: select samples". The URL is "pma.ipums.org/pma-action/samples". The page has a sidebar on the left with "Documentation" and a main content area. In the main area, there are two sections: "Sample Members" and "SUBMIT SAMPLE SELECTIONS". A red oval highlights the "Sample Members" section, which contains four radio button options: "Female Respondents" (selected), "Female Respondents and Household Members", "Female Respondents and Female Non-respondents", and "All Cases (Respondents and Non-respondents to Household and Female Questionnaires)". Below the "Sample Members" section is a "SUBMIT SAMPLE SELECTIONS" button. At the bottom of the page, there is a footer with "SUPPORTED BY: THE BILL & MELINDA GATES FOUNDATION, PMA, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA." and "COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA".

**Female Respondents** only includes women who completed *all or part* of a Female Questionnaire. This option selects all members of the panel study. In addition, it includes women who only participated in only one phase - we will demonstrate how to identify and drop these cases below.<sup>1</sup>

**Female Respondents and Female Non-respondents** includes all women who were eligible to participate in a Female Questionnaire. Eligible women are those age 15-49 who were listed on the roster collected in a Household Questionnaire. If an eligible woman declined the Female Questionnaire or was not available, variables associated with that questionnaire will be coded “Not interviewed (female questionnaire)”.

**PANELWOMAN** indicates whether an individual is a member of the panel study.

**RESULTFQ** indicates whether an individual completed the Female Questionnaire.

<sup>1</sup>Women who completed all or part of the Female Questionnaire in *more than one phase* of the study are considered **panel members**. Women who completed it only at Phase 1 are included in a longitudinal extract, but they are not **panel members**. Likewise, women who completed it for the first time at Phase 2 are included, but are not **panel members** if they 1) will reach age 50 before Phase 3, or 2) declined the invitation to participate again in Phase 3.

**Female Respondents and Household Members** adds records for all other members of a Female Respondent's household. These household members did not complete the Female Questionnaire, but were listed on the household roster provided by the respondent to a Household Questionnaire. Basic demographic variables are available for each household member, as are common wealth, water, sanitation, and other variables shared for all members of the same household.

**All Cases** includes all members listed on the household roster from a Household Questionnaire. If the Household Questionnaire was declined or if no respondent was available, any panel member appearing in other phases of the study will be coded "Not interviewed (household questionnaire)" for variables associated with the missing Household Questionnaire.

After you've selected samples and sample members for your extract, click the "Submit Sample Selections" button to return to the main data browsing menu.

ELIGIBLE indicates whether an individual was eligible for the female questionnaire.

RESULTHQ indicates whether a member of the individual's household completed the Household Questionnaire.

## 1.2 VARIABLE SELECTION

You can browse IPUMS PMA variables by topic or alphabetically by name, or you can [search](#) for a particular term in a variable name, label, value labels, or description.

The screenshot shows the IPUMS PMA website's variable selection page. At the top, there is a navigation bar with links for 'LOG IN | REGISTER | GLOBAL HEALTH | IPUMS.ORG'. On the right, a 'DATA CART' section shows '0 VARIABLES' and '6 SAMPLES' with a 'VIEW CART' button. The main content area has a header 'CURRENTLY BROWSING: "FAMILY PLANNING - PERSON"' with a 'CHANGE' link. Below this is a 'SELECT VARIABLES' section with dropdown menus for 'TOPICS' (set to 'TECHNICAL'), 'A-Z' (set to 'FAMILY PLANNING'), and 'SEARCH' (with a magnifying glass icon). To the right of these are 'DISPLAY OPTIONS' and 'HELP' and 'COUNTRY ABBREVIATIONS' links. A sidebar on the left lists various topics: 'SAMPLES' (selected), 'TECHNICAL', 'DEMOGRAPHICS (WOMEN)', 'FAMILY PLANNING' (selected), 'HEALTH', 'ABORTION', 'HOUSEHOLD', 'WATER AND SANITATION', and 'COVID-19'. The 'FAMILY PLANNING' topic is expanded, showing sub-topics: 'FERTILITY PREFERENCES', 'SEXUAL BEHAVIOR', 'CURRENT OR RECENT FAMILY PLANNING USE', 'PREVIOUS FAMILY PLANNING USE', 'EVER OR FIRST USE OF FAMILY PLANNING', 'DISCONTINUATION OF FAMILY PLANNING', 'NOT USING FAMILY PLANNING', 'FUTURE FAMILY PLANNING USE', 'FAMILY PLANNING ADVERTISEMENT', 'FAMILY PLANNING KNOWLEDGE', 'FAMILY PLANNING ACCESS', 'ATTITUDE TOWARDS FAMILY PLANNING', 'INFLUENCES ON FP', 'CONTRACEPTIVE ACCEPTABILITY', and 'CONTRACEPTIVE CALENDAR'. The bottom of the page includes a 'COPYRIGHT' section and a footer with 'MINNESOTA' and 'UNIVERSITY OF MINNESOTA' links. A small 'javascript:void(0);' is at the very bottom.

In this example, we'll select the [Discontinuation of Family Planning](#) topic. The availability of each associated variable is shown in a table containing all of the samples we've selected.

- x indicates that the variable is available for *all phases*
- / indicates that the variable is available for *one phase*
- – indicates that the variable is not available for *any phase*

You can click the + button to add a variable to your cart, or click a variable name to learn more.

The screenshot shows the IPUMS PMA website interface. At the top, there's a navigation bar with links for LOG IN, REGISTER, GLOBAL HEALTH, and IPUMS.ORG. On the right, a "DATA CART" section shows 0 VARIABLES and 6 SAMPLES, with a "VIEW CART" button. Below the navigation, the IPUMS PMA logo is displayed. The main content area has a header "CURRENTLY BROWSING: 'FAMILY PLANNING - PERSON'" with a "CHANGE" link. There are three tabs: "SELECT VARIABLES" (highlighted), "DISPLAY OPTIONS", and "HELP". Under "SELECT VARIABLES", there are buttons for "TOPICS", "A-Z", and "SEARCH". A note says "AN 'X' INDICATES THE VARIABLE IS AVAILABLE IN THAT DATASET." The main table is titled "DISCONTINUATION OF FAMILY PLANNING VARIABLES (TOP)" and "LONGITUDINAL SAMPLES". It lists variables like FPSTOPMO, EPIMPREMOVEYR, and EPIMPRMYYUNAVAIL, along with their labels and availability across datasets: BURKF, CONDR, CONDR, KENYA, NIGERA, and NIGERA. The table includes columns for "Add to cart" (with a plus sign icon) and "Variable Label". The bottom of the page includes a footer with copyright information: "SUPPORTED BY: THE BILL & MELINDA GATES FOUNDATION, PMA, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA." and "COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA."

DISCONTINUATION OF FAMILY PLANNING VARIABLES (TOP)		LONGITUDINAL SAMPLES							
Add to cart	Variable	Variable Label	Type	BURKF 2020 - 2021	CONDR 2019a - 2020a	CONDR 2019b - 2020b	KENYA 2019 - 2020	NIGERA 2019a - 2020a	NIGERA 2019b - 2020b
+ FPSTOPMO	Month stopped using most recent method	P X / / X . X X							
+ FPSTOPYR	Year stopped using most recent method	P X X X . X X							
+ FPSTOPUSECMC	Date stopped using recent method of FP in century month	P X X X . X X							
+ FPIMPREMOVEYR	Tried to remove implant in past 12 months	P X / / X / /							
+ EPIMPRMTRYLOC	Location of implant removal attempt	P X / / / /							
+ EPIMPRMYYCOST	Why implant not removed: Service cost	P X / / X / /							
+ EPIMPRMYYCOUND	Why implant not removed: Provider counseled against	P X / / / X / /							
+ EPIMPRMYYCLOSED	Why implant not removed: Facility closed	P X / / / X / /							
+ EPIMPRMYYOTH	Why implant not removed: Other	P X / / X / /							
+ EPIMPRMYYREFUSE	Why implant not removed: Provider refused	P X / / X / /							
+ EPIMPRMYYELSEWH	Why implant not removed: Referred elsewhere	P X / / X / /							
+ EPIMPRMYYRETURN	Why implant not removed: Told to return another day	P X / / X / /							
+ EPIMPRMYYTRAVEL	Why implant not removed: Travel cost	P X / / X / /							
+ EPIMPRMYYUNAVAIL	Why implant not removed: Qualified provider not available	P X / / X / /							
+ EPIMPRMYYUNSUCC	Why implant not removed: Failed attempt by provider	P X / / X / /							

## 1.2.1 Codes

Let's take a look at the variable PREGNANT. You'll find the variable name and label shown at the top of the page. Below, you'll see several tabs beginning with the CODES tab. For discrete variables, this tab shows all of the available codes and value labels associated with each response. You'll also see the same x, /, and – symbols in a table indicating the availability of each response in each sample.

“Case-count view” is not available for longitudinal samples, where each sample includes data from multiple phases. For cross-sectional samples, this option shows the frequency of each response.

The screenshot shows the IPUMS PMA website interface. At the top, there is a navigation bar with links for LOG IN, REGISTER, GLOBAL HEALTH, and IPUMS.ORG. On the right, there is a 'DATA CART' section showing 0 VARIABLES and 6 SAMPLES, with a 'VIEW CART' button. The main content area is titled 'PREGNANT' and shows the variable as 'Pregnancy status'. It includes a 'Group' section for 'Core demographics'. Below this, there is a table with tabs for CODES, DESCRIPTION, COMPARABILITY, UNIVERSE, AVAILABILITY, and QUESTIONNAIRE TEXT. The 'CODES' tab is selected. Under 'Codes and Frequencies', there are two radio button options: 'Category availability view' (selected) and 'Case-count view (Unavailable for longitudinal samples)'. A legend indicates that an 'X' means the category is available for that sample. The 'Female Respondents' option is also selected. The table below shows availability across six longitudinal samples: BURKF, CONDR, KENYA, NIGERA, and NIGERA 19b-20b. The table includes the following data:

		BURKF	COND R	KENYA	NIGERA	NIGERA 19b - 20b	
Code	Label	20 - 21	19a - 20a	19b - 20b	19 - 20	19a - 20a	19b - 20b
00	No	X	X	X	X	X	X
01	Yes	X	X	X	X	X	X
95	Not interviewed (female questionnaire)	.	.	.	.	.	.
96	Not interviewed (household questionnaire)	.	.	.	.	.	.
97	Don't know	X	X	X	X	X	X
98	No response	X	/	.	X	X	X
99	NIU (not in universe) or missing	.	.	.	.	.	.

At the bottom of the page, there is a footer with the text 'SUPPORTED BY: THE BILL & MELINDA GATES FOUNDATION, PMA, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA.' and 'COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA.'

Above, there are no responses for “Not interviewed (female questionnaire)” and “Not interviewed (household questionnaire)”; this is because only samples members included in a “Female Respondents” extract are displayed by default. If we instead choose “All Cases”, this variable will include those response options because we’ll include every person listed on the household roster (even if the Household or Female Questionnaire was not completed).

**PREGNANT**

Pregnancy status  
Group: [Core demographics](#)

**CODES** **DESCRIPTION** **COMPARABILITY** **UNIVERSE** **AVAILABILITY** **QUESTIONNAIRE TEXT**

**Codes and Frequencies**

Category availability view  
 Case-count view (Unavailable for longitudinal samples)

Female Respondents  
 Female Respondents and Household Members  
 Female Respondents and Female Non-respondents  
 All Cases (Respondents and Non-respondents to Household and Female Questionnaires)

An 'X' indicates the category is available for that sample

LONGITUDINAL SAMPLES						
Code	Label	BURKF	COND1	COND2	KENYA	NIGERA
		20 - 21	19a - 20a	19b - 20b	19 - 20	19a - 20a
00	No	X	X	X	X	X
01	Yes	X	X	X	X	X
95	Not interviewed (female questionnaire)	X	X	X	X	X
96	Not interviewed (household questionnaire)	X	X	X	X	X
97	Don't know	X	X	X	X	X
98	No response	X	/	:	X	X
99	NIU (not in universe) or missing	X	X	X	X	X

SUPPORTED BY: THE BILL & MELINDA GATES FOUNDATION, PMA, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA.

COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA.

The symbol / again indicates that a particular response is available for some - but not all - phases of the study. For PREGNANCY it indicates that one of the options was either unavailable or was not selected by any sample respondents in a particular phase. If a variable was not included in all phases of the study, all response options will be marked with this symbol. For example, consider the variable [COVIDCONCERN](#), indicating the respondent's level of concern about becoming infected with COVID-19.

The screenshot shows the IPUMS PMA website interface. At the top, there is a navigation bar with links for LOG IN | REGISTER | GLOBAL HEALTH | IPUMS.ORG and a 'Guest' status indicator. The main header features the IPUMS PMA logo and the text 'PERFORMANCE MONITORING FOR ACTION'. Below the header, there are links for HOME | SELECT DATA | MY DATA | SUPPORT. On the right side, there is a 'DATA CART' section showing '0 VARIABLES' and '6 SAMPLES' with a 'VIEW CART' button. The main content area is titled 'COVIDCONCERN' and includes 'ADD TO CART' and 'CHANGE SAMPLES' buttons. A sub-section titled 'Codes and Frequencies' is displayed, showing a table of codes and their labels across six longitudinal samples: BURKF, CONDR, CONDR, KENYA, NIGERA, and NIGERA. The table includes columns for 'Code' and 'Label', and rows for various response categories like 'Not concerned', 'A little concerned', etc. The table uses a grid system where some cells contain a diagonal slash (/) indicating limited availability across phases. At the bottom of the page, there is a footer with copyright information: 'SUPPORTED BY: THE BILL & MELINDA GATES FOUNDATION, PMA, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA.' and 'COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA.'

Because Phase 1 questionnaires were administered prior to the emergence of COVID-19, this variable only appeared on Phase 2 questionnaires. The symbol / indicates limited availability across phases.

## 1.2.2 Variable Description

You'll find a detailed description for each variable on the DESCRIPTION tab. This tab also indicates whether a particular question appeared on the Household or Female Questionnaire.

The screenshot shows a web browser window for the IPUMS PMA website. The URL is pma.ipums.org/pma-action/variables/PREGNANT#description\_section. The page title is "IPUMS PMA: descr: PREGNANT". The header includes the IPUMS PMA logo, navigation links for LOG IN | REGISTER | GLOBAL HEALTH | IPUMS.ORG, and a DATA CART section showing 0 VARIABLES and 6 SAMPLES with a VIEW CART button. The main content area is titled "PREGNANT" and describes it as a "Pregnancy status". It includes an "ADD TO CART" button and a "CHANGE SAMPLES" button. Below these are tabs for CODES, DESCRIPTION (which is selected), COMPARABILITY, UNIVERSE, AVAILABILITY, and QUESTIONNAIRE TEXT. The DESCRIPTION tab contains the text: "PREGNANT indicates whether or not the woman was pregnant at the time of the interview. The question associated with this variable was included in the female questionnaire." At the bottom of the page, there is a footer with links to SUPPORTED BY: THE BILL & MELINDA GATES FOUNDATION, PMA, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA, and a COPYRIGHT notice for MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA.

### 1.2.3 Comparability Notes

The COMPARABILITY tab describes important differences between samples. Additionally, it may contain information about similar variables appearing in DHS samples provided by IPUMS DHS.

The screenshot shows a web browser window for the IPUMS PMA website. The URL is pma.ipums.org/pma-action/variables/PREGNANT#comparability\_section. The page title is "IPUMS PMA: desc: PREGNANT". The top navigation bar includes links for LOGIN, REGISTER, GLOBAL HEALTH, and IPUMS.ORG. A "DATA CART" section on the right shows 0 VARIABLES and 6 SAMPLES with a "VIEW CART" button. The main content area is titled "PREGNANT" and shows "Pregnancy status" under "Group: Core demographics". Below this, there are tabs for CODES, DESCRIPTION, COMPARABILITY, UNIVERSE, AVAILABILITY, and QUESTIONNAIRE TEXT. The "COMPARABILITY" tab is selected, displaying the heading "Comparability" and the text: "There are minor universe differences among samples; see the Universe tab for more details." It also contains a section titled "Comparability with IPUMS-DHS" which states: "PREGNANT in IPUMS-PMA is similar to the variable PREGNANT in IPUMS-DHS. There may be differences in questionnaire text or the variable's universe; see the Survey Text and Universe Tab of the IPUMS-DHS variable for more information." At the bottom of the page, there is a footer note: "SUPPORTED BY: THE BILL & MELINDA GATES FOUNDATION, PMA, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA." and a copyright notice: "COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA."

## 1.2.4 Sample Universe

The UNIVERSE tab describes selection criteria for this question. In this case, there are some differences between samples:

- In DRC samples, all women aged 15-49 received this question.
- For all other samples, the question was skipped if any such woman previously indicated that she was menopausal or had a hysterectomy.

The screenshot shows the IPUMS PMA website interface. At the top, the IPUMS PMA logo is visible along with navigation links for LOGIN, REGISTER, GLOBAL HEALTH, and IPUMS.ORG. A 'DATA CART' section indicates 0 VARIABLES and 6 SAMPLES, with a 'VIEW CART' button. The main content area is titled 'PREGNANT' with 'ADD TO CART' and 'CHANGE SAMPLES' buttons. Below this, a tab menu includes 'CODES', 'DESCRIPTION', 'COMPARABILITY', 'UNIVERSE' (which is selected), 'AVAILABILITY', and 'QUESTIONNAIRE TEXT'. The 'UNIVERSE' tab displays a list of survey descriptions from various countries and years, each specifying age (15-49) and status (not menopausal, not had a hysterectomy). The bottom of the page includes support information and copyright details.

SUPPORTED BY: THE BILL & MELINDA GATES FOUNDATION, PMA, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA.

COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA.

## 1.2.5 Availability Across Samples

The AVAILABILITY tab shows all other samples (including cross-sectional samples) where this variable is available.

The screenshot shows the IPUMS PMA website interface. At the top, there is a navigation bar with links for LOG IN, REGISTER, GLOBAL HEALTH, and IPUMS.ORG. On the right side of the header, there is a "DATA CART" section indicating 0 VARIABLES and 6 SAMPLES, with a "VIEW CART" button. Below the header, the IPUMS PMA logo is displayed, along with the text "PERFORMANCE MONITORING FOR ACTION", "HOME | SELECT DATA | MY DATA | SUPPORT", and a "PREGNANT" category. Under the "PREGNANT" category, there are two buttons: "ADD TO CART" and "CHANGE SAMPLES". Below these buttons, it says "Pregnancy status" and "Group: Core demographics". A horizontal navigation bar below the status includes tabs for CODES, DESCRIPTION, COMPARABILITY, UNIVERSE, AVAILABILITY (which is highlighted in purple), and QUESTIONNAIRE TEXT. The main content area is titled "Availability" and lists the following countries and years: Burkina Faso: 2014-2018, 2020-2021; Congo (Democratic Republic): 2013-2020; Cote d'Ivoire: 2017-2018, 2020; Ethiopia: 2014-2019; Ghana: 2013-2017; India: 2016-2018, 2020; Indonesia: 2015-2016; Kenya: 2014-2020; Niger: 2015-2018; Nigeria: 2014-2020; Uganda: 2014-2020. At the bottom of the page, there is a footer with the text "SUPPORTED BY: THE BILL & MELINDA GATES FOUNDATION, PMA, STAT/TRANSFER, AND UNIVERSITY OF MINNESOTA." and "COPYRIGHT © MINNESOTA POPULATION CENTER, UNIVERSITY OF MINNESOTA."

## 1.2.6 Questionnaire Text

Finally, you'll find the full text of each question on the **QUESTIONNAIRE TEXT** tab. Each phase of the survey is shown separately, and you may click the "view entire document: text" link to view the complete questionnaire for a particular sample in any given phase.

The screenshot shows a web browser window for the IPUMS PMA website. The URL is [pma.ipums.org/pma-action/variables/PREGNANT#questionnaire\\_text\\_section](https://pma.ipums.org/pma-action/variables/PREGNANT#questionnaire_text_section). The page title is "IPUMS PMA: descr: PREGNANT". The top navigation bar includes links for LOG IN | REGISTER | GLOBAL HEALTH | IPUMS.ORG and a Guest account. A "DATA CART" section on the right shows "YOUR DATA EXTRACT", "0 VARIABLES", and "6 SAMPLES" with a "VIEW CART" button. The main content area is titled "PREGNANT" and shows "Pregnancy status" under the "Group: Core demographics" heading. Below this, there are tabs for CODES, DESCRIPTION, COMPARABILITY, UNIVERSE, AVAILABILITY, and QUESTIONNAIRE TEXT. The QUESTIONNAIRE TEXT tab is selected and displays the following content:

**Questionnaire Text**

Country	Year	Location
Burkina Faso	2020	Congo (Democratic Republic) Kenya 2019 2019b
Burkina Faso	2021	Congo (Democratic Republic) Kenya 2020 2020a
Congo (Democratic Republic)	2019a	Congo (Democratic Republic) Nigeria 2019a 2020b

[Burkina Faso 2020](#) [top](#)

[Questionnaire form](#) [view entire document: text](#)

14. Are you pregnant now?

Yes  
 No  
 Unsure  
 No response

[Burkina Faso 2021](#) [top](#)

[Questionnaire form](#) [view entire document: text](#)

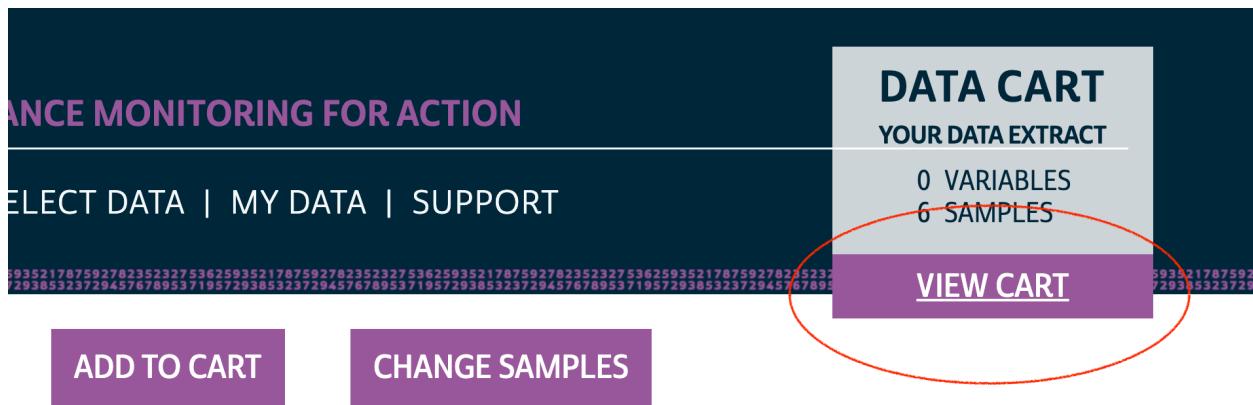
14. Are you pregnant now?

Yes  
 No  
 Unsure  
 No response

[Congo \(Democratic Republic\) 2019a](#) [top](#)

## 1.2.7 Checkout

Use the buttons at the top of this page to add the variable to your Data Cart, or to “VIEW CART” and begin checkout.



## 1.3 DATA FOR R USERS

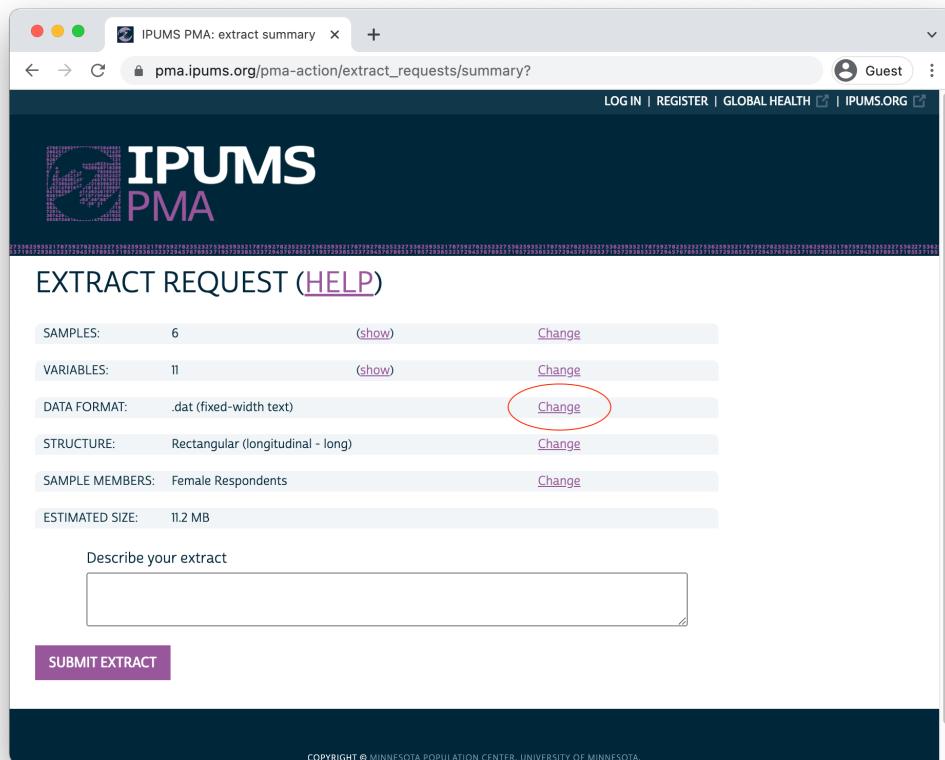
Your Data Cart shows all of the variables you've selected, plus several “preselected” variables that will be automatically included in your extract. Click the “CREATE DATA EXTRACT” button to prepare your download.

The screenshot shows the IPUMS PMA Data Cart interface. At the top right, there's a "DATA CART" section indicating "1 VARIABLE" and "6 SAMPLES". Below this, the main title is "DATA CART". There are three buttons: "ADD MORE VARIABLES", "CREATE DATA EXTRACT" (which is highlighted with a red oval), and "ADD MORE SAMPLES". A "Clear Data Cart" link is also present. The main area displays a table of variables and their types across different countries and years. The table includes columns for Type (P or X), and rows for variables like SAMPLE, COUNTRY, YEAR, ELIGIBLE, EAID, CONSENTFO, FQINSTITD, CONSENTHQ, FOWWEIGHT, STRATA, and PANELWOMAN. The "CREATE DATA EXTRACT" button is located at the bottom left of the main content area.

In cart	Variable	Variable Label	Type	BURKF 2020 - 2021	CONDRI 2019a - 2020a	CONDRI 2019b - 2020b	KENYA 2019 - 2020	NIGERA 2019a - 2020a	NIGERA 2019b - 2020b
<input checked="" type="checkbox"/>	SAMPLE	PMA sample number [preselected]	P	X	X	X	X	X	X
<input checked="" type="checkbox"/>	COUNTRY	PMA country [preselected]	P	X	X	X	X	X	X
<input checked="" type="checkbox"/>	YEAR	Year [preselected]	P	X	X	X	X	X	X
<input checked="" type="checkbox"/>	ELIGIBLE	Eligible female respondent [preselected]	P	X	X	X	X	X	X
<input checked="" type="checkbox"/>	EAID	Enumeration area [preselected]	P	X	X	X	X	X	X
<input checked="" type="checkbox"/>	CONSENTFO	Female respondent provided consent to be interviewed [preselected]	P	X	X	X	X	X	X
<input checked="" type="checkbox"/>	FQINSTITD	Unique ID for female questionnaire [preselected]	P	X	X	X	X	X	X
<input checked="" type="checkbox"/>	CONSENTHQ	Household respondent provided consent to be interviewed [preselected]	P	X	X	X	X	X	X
<input checked="" type="checkbox"/>	FOWWEIGHT	Female weight [preselected]	P	X	X	X	X	X	X
<input checked="" type="checkbox"/>	STRATA	Strata [preselected]	P	X	.	.	X	X	X
<input checked="" type="checkbox"/>	PANELWOMAN	Panel woman interviewed in Phase 1	P	/	/	/	/	/	/

### 1.3.1 Select a Fixed-width File

Before you submit an extract request, you'll have the opportunity to choose a "Data Format". **R users should selected Fixed-width text (.dat)** - you'll notice that data formatted for Stata, SPSS, and SAS are also available. CSV files are provided, but not recommended. (If you wish to change Sample Members, you may do so again here.)



The screenshot shows a web browser window for the IPUMS PMA website. The title bar reads "IPUMS PMA: extract summary". The URL is "pma.ipums.org/pma-action/extract\_requests/summary?". The top navigation bar includes "LOG IN | REGISTER | GLOBAL HEALTH | IPUMS.ORG". The main content area is titled "EXTRACT REQUEST ([HELP](#))". It displays the following settings:

SAMPLES:	6	( <a href="#">show</a> )	<a href="#">Change</a>
VARIABLES:	11	( <a href="#">show</a> )	<a href="#">Change</a>
DATA FORMAT:	.dat (fixed-width text)		<a href="#">Change</a> (circled in red)
STRUCTURE:	Rectangular (longitudinal - long)		<a href="#">Change</a>
SAMPLE MEMBERS:	Female Respondents		<a href="#">Change</a>
ESTIMATED SIZE:	11.2 MB		

Below these settings is a text input field labeled "Describe your extract" with a placeholder text "Describe your extract". At the bottom is a purple "SUBMIT EXTRACT" button.

Once the Fixed-width option is selected, you may add a description and then proceed to the download page.

### 1.3.2 Download

After a few moments, you'll receive an email indicating that your extract has been created. You'll need to obtain two files from the download page:

- Click the green “Download DAT” button to download the data file. You’ll receive a file with a number like pma\_00001.dat.gz.
- Right click on “DDI” and click “Save link as”. You’ll receive a corresponding XML file like pma\_00001.xml.

The screenshot displays two tables of extracts on the IPUMS website. The top table lists extracts by date, with a green 'Download DAT' button and a red box around the 'DDI' link. The bottom table shows a context menu for the 'DDI' link, with 'Save link as...' highlighted in red. Arrows point from the numbered steps below to the respective buttons and menu item.

1) Click here to download the data.

2) Right click here to select the DDI.

3) Then select "Save link as..." (or "Download Linked File") to save the DDI.

Place both files in a folder that R can use as its [working directory](#). We **strongly recommend** using [RStudio projects](#) to manage all of the files and analysis scripts used for a particular research project. We'll place our files in a subfolder called “data” within our own RStudio project folder.

Open RStudio (or R) and load the packages [ipumsr](#) and [tidyverse](#). If you are not using an RStudio project, you will need to change your working directory to match the location of your downloaded files.

```
library(ipumsr)
library(tidyverse)
setwd("~/Downloads") # ONLY if not using an RStudio project (change as needed)
```

We'll now demonstrate loading both a long and a wide extract, and we'll take a brief look at the structure of each.

## 1.4 LONG DATA STRUCTURE

We've downloaded a **long** data extract (Female Respondents only) and saved it in a folder called "data" in our working directory. We'll now load it into R as an object called `long`.

To load an IPUMS PMA extract into R, you'll need to reference *both* the DDI file *and* the fixed-width data file in the function `read_ipums_micro` from `ipumsr`.

```
long <- read_ipums_micro(  
  ddi = "data/pma_00095.xml",  
  data = "data/pma_00095.dat.gz"  
)
```

In a **long** extract, data from each phase will be organized in *separate rows*. Here, responses from three panel members are shown:

```
long %>%  
  filter(FQINSTID %>% str_starts("011") | FQINSTID %>% str_starts("015")) %>%  
  arrange(FQINSTID) %>%  
  select(FQINSTID, PHASE, AGE, PANELWOMAN)
```

```
# A tibble: 6 × 4  
FQINSTID          PHASE      AGE PANELWOMAN  
<chr>            <int+lbl> <int+lbl> <int+lbl>  
1 011W5S0HN91I4H4I3T9JCMBHB 1 [Baseline]    29     NA  
2 011W5S0HN91I4H4I3T9JCMBHB 2 [First follow up] 30     1 [Yes]  
3 015NP6FJTIA98FYCBBBS1F0F7 1 [Baseline]    47     NA  
4 015NP6FJTIA98FYCBBBS1F0F7 2 [First follow up] 48     1 [Yes]  
5 015WYNN02WXHH6JA4HA9PL1MR 1 [Baseline]    20     NA  
6 015WYNN02WXHH6JA4HA9PL1MR 2 [First follow up] 21     1 [Yes]
```

Each panel member receives a unique ID shown in `FQINSTID`. The variable `PHASE` shows that each woman's responses to the Phase 1 Female Questionnaire appears in the first row, while her Phase 2 responses appear in the second. `AGE` shows each woman's age when she completed the Female Questionnaire for each phase.

`PANELWOMAN` indicates whether the woman completed all or part of the Female Questionnaire in a *prior* phase, and that she'd agreed to continue participating in the panel study at that time. The value `NA` appears in the rows for Phase 1, as `PANELWOMAN` was not included in Phase 1 surveys.

We mentioned above that you'll also include responses from some non-panel members when you request an extract with Female Respondents. These include women who did not complete all or part the Female Questionnaire in a prior phase, as indicated by [PANELWOMAN](#). These women are not assigned a value for [FQINSTID](#) - instead, you'll find an empty string:

```
long %>% count(PHASE, PANELWOMAN, FQINSTID == "")
```

```
# A tibble: 3 × 4
  PHASE      PANELWOMAN `FQINSTID == ""` n
  <int+lbl>    <int+lbl> <lgl>        <int>
1 1 [Baseline]     NA    FALSE        23591
2 2 [First follow up] 0 [No]   TRUE        6586
3 2 [First follow up] 1 [Yes]  FALSE       18194
```

For most longitudinal analysis applications, you'll need to drop non-panel members together with any women who did not fully complete the Phase 2 Female Questionnaire. We'll demonstrate using [group\\_by](#) to ensure that there is one row for every [FQINSTID](#) where [PHASE == 1](#) and another row where [PHASE == 2 & RESULTFQ == 1](#).

```
long <- long %>%
  group_by(FQINSTID) %>%
  filter(any(PHASE == 1) & any(PHASE == 2 & RESULTFQ == 1)) %>%
  ungroup()
```

The [PMA Longitudinal Briefs](#) published for each sample also include only members of the *de facto* population. These are women who slept in the household during the night prior to the interview for each Household Questionnaire, such that [RESIDENT](#) takes the value 11 or 22. We'll use [group\\_by](#) again to include only *de facto* women from both phases.

```
long <- long %>%
  group_by(FQINSTID) %>%
  filter(all(RESIDENT %in% c(11, 22))) %>%
  ungroup()
```

Following these steps, you can check the size of each analytic sample like so. (Reminder: samples for DRC and Nigeria are sub-nationally representative, so we'll show separate frequencies for each [GEOCD](#) and [GEONG](#)).

```
long %>% count(COUNTRY, GEOCD, GEONG, PHASE)
```

# A tibble: 12 × 5				
	COUNTRY	GEOCD	GEONG	PHASE
	<int+lbl>	<int+lbl>	<int+lbl>	<int>
1	1 [Burkina Faso]	NA	NA	1 [Baseline] 5212
2	1 [Burkina Faso]	NA	NA	2 [First follow u... 5212
3	2 [Congo, Democratic Republic]	1 [Kinshasa]	NA	1 [Baseline] 1973
4	2 [Congo, Democratic Republic]	1 [Kinshasa]	NA	2 [First follow u... 1973
5	2 [Congo, Democratic Republic]	2 [Kongo Central]	NA	1 [Baseline] 1514
6	2 [Congo, Democratic Republic]	2 [Kongo Central]	NA	2 [First follow u... 1514
7	7 [Kenya]	NA	NA	1 [Baseline] 6939
8	7 [Kenya]	NA	NA	2 [First follow u... 6939
9	9 [Nigeria]	NA	2 [Lagos]	1 [Baseline] 1089
10	9 [Nigeria]	NA	2 [Lagos]	2 [First follow u... 1089
11	9 [Nigeria]	NA	4 [Kano]	1 [Baseline] 998
12	9 [Nigeria]	NA	4 [Kano]	2 [First follow u... 998

## 1.5 WIDE DATA STRUCTURE

We've also downloaded a **wide** data extract (Female Respondents only) and saved it in the "data" folder in our working directory. We'll also load this extract into R as an object named `wide`.

```
wide <- read_ipums_micro(  
  ddi = "data/pma_00084.xml",  
  data = "data/pma_00084.dat.gz"  
)
```

In a **wide** extract, all of the responses from one woman appear in the *same row*. The IPUMS extract system appends a numeric suffix to each variable name corresponding with the phase from which it was drawn. Consider our three example panel members again:

```
wide %>%  
  filter(FQINSTID %>% str_starts("011") | FQINSTID %>% str_starts("015")) %>%  
  select(FQINSTID, AGE_1, AGE_2, PANELWOMAN_1, PANELWOMAN_2)
```

```
# A tibble: 3 × 5  
FQINSTID          AGE_1     AGE_2    PANELWOMAN_1 PANELWOMAN_2  
<chr>            <int+lbl> <int+lbl> <int+lbl>   <int+lbl>  
1 011W5S0HN91I4H4I3T9JCMBHB 29        30       NA         1 [Yes]  
2 015NP6FJTIA98FYCBBBS1F0F7 47        48       NA         1 [Yes]  
3 015WYNN02WXHH6JA4HA9PL1MR 20        21       NA         1 [Yes]
```

Each panel member has one unique ID shown in `FQINSTID`. However, `AGE` is parsed into two columns: `AGE_1` shows each woman's age at Phase 1, and `AGE_2` shows her age at Phase 2.

As we've discussed, `PANELWOMAN` is not available for Phase 1, as it indicates whether the woman completed all or part of the Female Questionnaire in a *prior* phase. For this reason, all values in `PANELWOMAN_1` are NA. Most variables are copied once for each phase, even if they - like `PANELWOMAN_1` - are not available for all phases.

You might expect the total length of a **wide** extract to be half the length of a corresponding **long** extract. This is not the case! A **wide** extract includes one row for each woman who completed all or part of the Female Questionnaire *for any phase* - you'll find placeholder columns for phases where the interview was not conducted.

```
wide %>%
  filter(FQINSTID == "0C8VQU6B03BXLAVVZ8SB90EKQ") %>%
  select(RESULTFQ_1, AGE_1, RESULTFQ_2, AGE_2)

# A tibble: 1 × 4
  RESULTFQ_1     AGE_1     RESULTFQ_2     AGE_2
  <int+lbl>     <int+lbl> <int+lbl>     <int+lbl>
1 1 [Completed] 31          2 [Not at home] 95 [Not interviewed (female questionnaire)]
```

In a **long** extract, rows for the missing phase are dropped. In this example, the woman was “not at home” for the Phase 2 Female Questionnaire. When we select a **long** extract containing only Female Respondents, her Phase 2 row is excluded automatically (it will be included if you request an extract containing Female Respondents and Female Non-respondents).

```
long %>%
  filter(FQINSTID == "0C8VQU6B03BXLAVVZ8SB90EKQ") %>%
  select(PHASE, RESULTFQ, AGE)

# A tibble: 1 × 3
  PHASE     RESULTFQ     AGE
  <int+lbl> <int+lbl> <int+lbl>
1 1 [Baseline] 1 [Completed] 31
```

Again: for most longitudinal analysis applications, you’ll need to remove cases where women were not interviewed for Phase 1 or where the Phase 2 Female Questionnaire was not completed:

```
wide <- wide %>% filter(RESULTFQ_2 == 1 & !is.na(RESULTFQ_1))
```

The *de facto* population appearing in [PMA Longitudinal Briefs](#) is defined in **wide** extracts by cases where the values 11 or 12 appear in *both* RESIDENT\_1 and RESIDENT\_2:

```
wide <- wide %>% filter(RESIDENT_1 %in% c(11, 22) & RESIDENT_2 %in% c(11, 22))
```

Following these steps, each analytic sample contains the same number of cases shown in the final **long** format extract above.

```
wide %>%
  group_by(COUNTRY, GEOCD, GEONG) %>%
  count()

# A tibble: 6 × 4
# Groups:   COUNTRY, GEOCD, GEONG [6]
  COUNTRY           GEOCD      GEONG     n
  <int+lbl>        <int+lbl>    <int+lbl> <int>
1 1 [Burkina Faso]    NA        NA     5212
2 2 [Congo, Democratic Republic] 1 [Kinshasa]    NA     1973
3 2 [Congo, Democratic Republic] 2 [Kongo Central] NA     1514
4 7 [Kenya]          NA        NA     6939
5 9 [Nigeria]        NA        2 [Lagos]  1089
6 9 [Nigeria]        NA        4 [Kano]   998
```

## 1.6 WHICH FORMAT IS BEST FOR ME?

The choice between **long** and **wide** formats ultimately depends on your research objectives.

Many data manipulation tasks, for example, are faster and easier to perform in the **wide** format. In the example above, we needed to identify women who completed a Female Questionnaire and were members of the *de facto* population in both phases. In the **long** format, we first had to group the data by `FQINSTID` with `group_by`, thereby ensuring that a Phase 1 and Phase 2 check could be performed for each woman. In preparing for this post, this approach took about 36.5 seconds. By comparison, the same task was achieved without `group_by` in **wide** format in just 0.16 seconds. If your workflow requires multiple comparisons between phases, the **wide** format may be the best choice!

On the other hand, many of the longitudinal modeling packages available for R require data to be in a **long** format - this includes both the [survival](#) package for Cox regression and the [lme4](#) package for multilevel models. Users who prefer the **wide** format for data cleaning and exploration can manually switch to **long** format with help from [pivot\\_longer](#), for example:

```
wide %>% select(FQINSTID, AGE_1, PREGNANT_1, AGE_2, PREGNANT_2)
```

FQINSTID	AGE_1	PREGNANT_1	AGE_2	PREGNANT_2
<chr>	<int+lbl>	<int+lbl>	<int+lbl>	<int+lbl>
1 uuid:0005f6d7-b7cd-46f6-8a6f-5f051b6ab4a2	30	0 [No]	31	0 [No]
2 uuid:0006cb76-09d1-4f2a-a92d-c12fcacf194b5	34	1 [Yes]	34	0 [No]
3 uuid:00204481-5cae-4188-abb3-0367d0ed9c14	17	0 [No]	18	0 [No]
4 uuid:002398f4-8f2d-4095-8019-c306d39cf2b9	29	0 [No]	29	0 [No]
5 uuid:00407300-c1e6-4e24-ab8d-8af5e1ca85a6	25	0 [No]	25	0 [No]
6 uuid:00413ed1-d176-44fb-a232-7e53c1db0958	32	0 [No]	32	0 [No]
7 uuid:0048a052-66ff-4ed5-9fa9-fc72e6dab696	38	0 [No]	39	0 [No]
8 uuid:004d80f0-90c6-4b77-bb4d-21d09c84fe74	38	0 [No]	38	0 [No]
9 uuid:00504cf5-870c-4a02-aad7-ea5d47b135ff	33	0 [No]	34	0 [No]
10 uuid:00534792-fb84-47b4-8606-e145d74cd089	24	0 [No]	25	0 [No]
11 uuid:0058ccb8-9892-4a60-b9ed-fb556a21f862	29	0 [No]	30	0 [No]
12 uuid:00682e93-0483-42b4-8f8d-2e0c36a26d37	16	0 [No]	17	0 [No]
13 uuid:007699e0-d078-4a30-ace2-6de2cfaa571a	29	0 [No]	30	0 [No]
14 uuid:00776552-ad8a-455b-93e0-ac9d066e2a85	22	0 [No]	23	0 [No]
15 uuid:0084caf8-e83f-45ee-9906-8a1d2d0b923a	28	0 [No]	29	0 [No]
16 uuid:009dbdcf-f9f0-4cb6-94e2-e6a46fb23bd6	41	0 [No]	42	0 [No]
17 uuid:00b89b1d-25e1-40f0-8e52-e06ce5525555	47	1 [Yes]	47	0 [No]
18 uuid:00bf8df9-340f-454b-95b7-04921ef54c28	15	0 [No]	16	0 [No]
19 uuid:00c8312a-561a-4855-840d-63a6fba578ae	22	0 [No]	22	0 [No]
20 uuid:00f7a6fb-484e-4664-a646-79dd480ab43e	24	0 [No]	25	0 [No]
# ... with 17,705 more rows				

With [pivot\\_longer](#), you can strip the suffix 1 or 2 from each variable, placing the result in a new column called PHASE. Then, we'll pivot each woman's age and pregnancy status from 2 **wide** columns into a single **long** one.

```
wide %>%
  select(FQINSTID, AGE_1, PREGNANT_1, AGE_2, PREGNANT_2) %>%
  pivot_longer(
    !FQINSTID,
    names_pattern = "(.*)_([1-2])",
    names_to = c(".value", "PHASE")
  )
```

```
# A tibble: 35,450 × 4
  FQINSTID          PHASE AGE PREGNANT
  <chr>            <chr> <int+lbl> <int+lbl>
1 uuid:0005f6d7-b7cd-46f6-8a6f-5f051b6ab4a2 1     30      0 [No]
2 uuid:0005f6d7-b7cd-46f6-8a6f-5f051b6ab4a2 2     31      0 [No]
3 uuid:0006cb76-09d1-4f2a-a92d-c12fcacf194b5 1     34      1 [Yes]
4 uuid:0006cb76-09d1-4f2a-a92d-c12fcacf194b5 2     34      0 [No]
5 uuid:00204481-5cae-4188-abb3-0367d0ed9c14 1     17      0 [No]
6 uuid:00204481-5cae-4188-abb3-0367d0ed9c14 2     18      0 [No]
7 uuid:002398f4-8f2d-4095-8019-c306d39cf2b9 1     29      0 [No]
8 uuid:002398f4-8f2d-4095-8019-c306d39cf2b9 2     29      0 [No]
9 uuid:00407300-c1e6-4e24-ab8d-8af5e1ca85a6 1     25      0 [No]
10 uuid:00407300-c1e6-4e24-ab8d-8af5e1ca85a6 2     25      0 [No]
11 uuid:00413ed1-d176-44fb-a232-7e53c1db0958 1     32      0 [No]
12 uuid:00413ed1-d176-44fb-a232-7e53c1db0958 2     32      0 [No]
13 uuid:0048a052-66ff-4ed5-9fa9-fc72e6dab696 1     38      0 [No]
14 uuid:0048a052-66ff-4ed5-9fa9-fc72e6dab696 2     39      0 [No]
15 uuid:004d80f0-90c6-4b77-bb4d-21d09c84fe74 1     38      0 [No]
16 uuid:004d80f0-90c6-4b77-bb4d-21d09c84fe74 2     38      0 [No]
17 uuid:00504cf5-870c-4a02-aad7-ea5d47b135ff 1     33      0 [No]
18 uuid:00504cf5-870c-4a02-aad7-ea5d47b135ff 2     34      0 [No]
19 uuid:00534792-fb84-47b4-8606-e145d74cd089 1     24      0 [No]
20 uuid:00534792-fb84-47b4-8606-e145d74cd089 2     25      0 [No]
# ... with 35,430 more rows
```

Manipulating patterns in variable names with [pivot\\_longer](#) takes practice, and we imagine many users will find it easier to simply work with data in the **long** format from the beginning.

Fortunately, the IPUMS PMA extract system makes it easy to select the samples, sample members, and variables that matter to your particular research question. Choices for **long** and **wide** data formats save an additional data cleaning step, allowing you to jump into longitudinal analysis as quickly as possible.