

4796138925900024972846961  
28625141000000000321437  
31547000000000000004598  
92870000000000000000121  
347000000000000000000634  
150000000000000000000018289  
67410000000000000000076566465  
50205217000000000000082352327  
795729385300000000000457678953  
0047386450000000000002165983721  
325214791600000000000181421539890  
941982980000000000000413634619730  
836160000000000000000451573948404  
19200000000000000000006396588002  
680000000000000000000100362100087  
56300000000000000000000719  
729100000000000000000002642  
38742900000000000000000431925  
59387246190000000000001479254386

# IPUMS

## PMA

### Expanding Opportunities For Longitudinal Research In Global Health And Population Studies: Analyzing Three-Phase Longitudinal Data In IPUMS PMA

October 24, 2023

#### Questions and Answers

*The following are the questions received during the live webinar and their answers. For additional questions or clarifications, contact IPUMS User Support at [ipums@umn.edu](mailto:ipums@umn.edu)*

#### What is the best way to deal with missing data? Drop the variable or the missing categories?

This depends on your analytical plan, as there are statistical debates on how to best handle missing data (e.g., listwise deletion, multiple imputation). IPUMS PMA does not provide any recommendation on how to handle these missing data, although we do provide as much detail as possible on the type of missing data for each variable (e.g., NIU, refused to answer, didn't know the answer).

#### What is a female non-respondent?

A female non-respondent is a woman who was eligible to be interviewed for the female survey, that is, if she was aged 15-49 and a resident of or visiting a sampled household, but she was not interviewed. Reasons for not being interviewed could include whether she refused to participate, was incapacitated, or was not available.

#### What is the missing % in current PMA, regarding this longitudinal part of the data?

This largely depends on the variable as the PMA questionnaires used precise skip patterns, so there is no short answer to your question.

Among the longitudinal samples for Burkina Faso, Democratic Republic of the Congo, Kenya, and Nigeria, between 60 to 70 percent of women initially interviewed in Phase 1 were also interviewed in both Phase 2 and Phase 3, as a result of attrition ranging from 12 to 27 percent of women between phases. See also slide 28 of our presentation.

#### Do you have a sample of questionnaires used for these surveys?

Yes, the questionnaires from the longitudinal panel can be found on the [questionnaire page](#) of the IPUMS PMA website. PDF format questionnaires are available in English as well as French

in selected countries. You can also see the questionnaire text associated with each variable on the Questionnaire Text tab.

### **Can I use the data to write a proposal and apply for a grant?**

If you are a registered user of IPUMS PMA, which is free, you may use the data for non-commercial purposes as long as you follow the data sharing agreement specified during the registration process. This includes drawing research proposals based on IPUMS PMA data.

### **For weighting, what if our reservation unit used for analysis is at the female level, but we also use the household level variables in our model, what should we do?**

We recommend that you use female level weights (either cross-sectional or longitudinal) even if you include household level characteristics in your model, because the household characteristic is attached to the female record.

### **How can we access data from multiple countries?**

In our [online data dissemination system](#), you can check the box for multiple countries and multiple years in the same custom data file on our “Select Samples” page.

### **Would there be a scenario that there are two variables, one is from female level and the other is from household level? If so, which weight variable (FQweight, household weight) should we use?**

The key to deciding which weight variable to use is the subpopulation you are calculating an estimate for. For example, if you are looking for the proportion of women who live in households with an improved source of drinking water, you would still use the female-level weight variable even though the variable originally comes from the household questionnaire.

### **Is there any way in the variable naming that we know certain variables are country-specific?**

In some cases, we include an abbreviation of the country name in the variable names of those variables that are country-specific, for example, GEOKE (1<sup>st</sup> level administrative geographic unit for Kenya). However, this is not true of all country-specific variables. Because of the harmonization work that IPUMS does with PMA data, you can depend on codes and variable names having consistent meanings across countries and years. Comparability issues will be documented on our Comparability tab for each variable.

### **Could you talk a little more about the overlapping contraceptive calendar data and how to extract them?**

In Phase 1, each female respondent who was pregnant or started using their current family planning method, given birth, or had a pregnancy end within the past 3 years was asked to provide month by month information about their contraceptive and fertility status in the 3 years prior to the interview. One year later, respondents were asked to provide the same information for the past 2 years prior to the interview, so there would be approximately one year of overlap between the Phase 1 and Phase 2 calendar data.

The calendar data in the original PMA data are formatted as a comma-delimited string listing the fertility or contraceptive status for each month prior to the interview date. The typical way to use

the calendar data is to use a command in your statistical software to split the string into separate variables representing each month in the recall period. We provide sample Stata code on our [contraceptive calendar user note](#). You can download the calendar data as it was formatted in the original PMA data from our website on the [contraceptive calendar variable group](#). We separate the calendar by country due to different interview months and recall periods.

**Do the data allow investigating events occurring between surveys? Does the data also collect some retrospective information like “whether a family member died in between surveys? If so, what date/year”?**

It is possible to use the contraceptive calendar data to identify women who had pregnancies, births, or contraceptive use episodes between phases that would not be captured by asking whether she was currently pregnant or currently using a method at the time of the survey. For these events, you would be able to identify the month and year. The questionnaires did not ask questions related to family member death, specifically. You could use the displaced GPS coordinates of the primary sampling unit to connect external data to the women’s records, such as natural disasters or other large-scale events.

**The data on non-respondent household members seems very limited in detail. What are suggested purposes for using it?**

One suggestion we have for a way to use the non-respondent household members is to use the roster to categorize or describe the household composition, and perhaps identify correlations between changes in household composition and the female respondent’s fertility or reproductive health outcomes.

**Can one analyze and publish this data in a peer review journal, if yes, do you provide IRBs used for the surveys if needed?**

Yes, we absolutely encourage the use of these data for publication in peer-reviewed journals. The data collectors with PMA have already been through the IRB process, so you may use the PMA data through us as secondary data without going through an IRB process yourself.

**For longitudinal household-level analysis, if we want to draw conclusions at the household level, which weight should we use? Are the panel weights that you discussed in the slides only for female-level analysis?**

Because of the sample design of the PMA longitudinal data, it is not recommended that you attempt to conduct a longitudinal analysis at the household level, as there are no household-level longitudinal weights. The panel weights discussed in the webinar are only at the female level.

**Do you require manuscripts written based on analyses of the IPUMS data to be shared with you prior to publishing?**

We do not require manuscripts to be shared with us before publishing, but we do ask that you cite us in your references. Please see our [citation page](#) for a suggested citation of IPUMS PMA data.

**Do the data have information about discontinuation of contraceptive options?**

Yes, both the cross-sectional and longitudinal data have information about contraceptive discontinuation. The contraceptive calendar data also include a variable that indicates the reason for discontinuation of each contraceptive use episode.

**Is the data only at national level or including subnational levels?**

Most PMA samples are representative at the national level with detail at the 1<sup>st</sup> administrative geographic level. Some samples are representative only at certain subnational geographies such as Rajasthan, India, or Kinshasa, DRC. In the case of Rajasthan, there is geographic detail at the 2<sup>nd</sup> administrative geographic level.

**Has PMA noted any particular selection in the attrition patterns of respondents?**

We are not aware of any studies on patterns of attrition for the PMA panel data, but it is possible to use the data to identify patterns because non-respondents are included in the data. We encourage you to follow publications by the data collectors at [pmapdata.org](http://pmapdata.org) in the event that they publish reports relevant to this topic.