User Notes for PMA-Ethiopia Panel Cohort 2 (6-week Follow-up) Survey Dataset, Version 1.0

Disclaimer: PMA cannot provide in-depth support for data analysis or data related questions, however, to assist the end-user, explanation of some variables is provided below.

Note: This is not a complete 6-week follow-up dataset. Users will need to combine this dataset with the Panel Cohort 2 Baseline dataset to get a complete 6-week follow-up dataset. See below for more information.

PMA Ethiopia

Performance Monitoring for Action Ethiopia (PMA Ethiopia) builds on the previous success of PMA2020/Ethiopia and PMA Maternal and Newborn Health study in the Southern Nations, Nationalities and Peoples Region (SNNP). PMA Ethiopia is a five-year project implemented in collaboration with Addis Ababa University, Johns Hopkins University, and the Federal Ministry of Health. It measures key reproductive, maternal and newborn health (RMNH) indicators.

Cross-sectional data, including a health facility-based survey, are collected annually in all regions. Longitudinal data (following pregnant women through one year postpartum) are collected in two cohorts of women (2019-2021 and 2021-2023) in four large, predominantly agrarian regions: Tigray, Oromiya, Amhara, and Southern Nations, Nationalities, and Peoples' Region, and one urban region, Addis Ababa. Afar is included in the first cohort (2019-2021) of the longitudinal survey.

Sampling

PMAET Panel Cohort 2 6-week Follow-up survey followed women who were enrolled into the Panel Cohort 2 Baseline Survey. Women who were 5-9 weeks postpartum at baseline who consented to the baseline survey, women 0-4 weeks postpartum at baseline who consented to the baseline and consented to follow-up, and women pregnant at baseline who consented to the baseline and consented to follow-up were eligible for the 6-week follow-up survey.

Women (n=273) who were 5-9 weeks postpartum at baseline received the 6-week follow-up survey at the same time as their baseline interview, conducted between November 2021 and January 2022. Women (n=2,024) 0-4 weeks postpartum or pregnant at baseline received a 6-week follow-up survey approximately 6-week after the delivery. A total of 2,072 women completed the 6-weeks follow-up survey conducted between November 2021 and October 2022.

For more information on the PMA survey methodology and sampling, please refer to PMA Survey Methodology at https://www.pmadata.org/data/survey-methodology.

Materials included in with this data zipped file

The data zipped file includes:

1. PMAET_Panel_Cohort2_6wkFU_v#_Date (Dataset in 3 file formats)
2. PMAET_Panel_Cohort2_6wkFU_FemaleQuestionnaire_v#_Date (pdf)
3. PMAET_Panel_Cohort2_6wkFU_Usernotes_v#_Date (pdf)
4. Example1_GenerateFull6wkFU_Append_v# (Stata dofile)
5. Example2_Merge1_BL&6wkFU_v# (Stata dofile)
6. Example3_Merge2_BL&Full6wkFU_v#_Date (Stata dofile)

Codebook

The latest version of the PMA Ethiopia Household and Female Survey master codebook can be downloaded from the https://www.pmadata.org/data-codebooks.

Six-week Follow-up Questionnaire

In the 6-week follow-up questionnaire, some groups of questions were repeated for each event (i.e., the same set of questions were repeated for each pregnancy or for each baby). The data for the repeat group was transformed to wide, and each variable name would have a root with a number starting with 1 (eg: variable names starting with **birth1**, **birth2**, **baby1**, and **baby2**, etc…).

Reproductive Coercion Questions:

All 6-week follow-up eligible women were asked the questions on reproductive coercion and intimate partner violence (IPV).

Antenatal Care (ANC) & Postnatal Care (PNC) Questions:

All 6-week eligible women received the questions on antenatal care (ANC), delivery, birth outcomes, neonatal care, and postnatal care (PNC).

General Variables

SIF variables: Date and time variables are provided in both string format and as Stata Internal Format (SIF) values. The variable name of any variable that has been changed into SIF is appended with SIF (e.g. **system_date** and **system_dateSIF**). The *"Do not Know"* value for dates is *Jan 1, 2030*. For each date question where the woman did not know the month but knew the year, the value is Jan 1 of that year and the variable with the same name as the date variable but ending in **_m_dnk** will have a value of 1.

Select multiple variables: Some questions allow for the selection of multiple answers. These variables are in string format and the values are the concatenation of answer choices. Multi-select options are generally, though not always, transformed into binary variables for analysis.

Variable Response Options

Select one: For the select one numeric variables, consistent values for option choices were given across PMA countries. Some select one variables, however, such as **region**, have answer options that vary across countries.

Select multiple: Similarly, most select multiple variables have the same response options across all PMA countries.

See the PMAET HQFQ Master Codebook for complete details on variables and answer choices for each survey.

Specific Variables

EA_ID: The primary sampling unit masked with a random number for anonymity. The same random number is applied to the same EA across multiple survey years.

RE_ID: Identification number of the resident enumerator (RE), or interviewer. RE names are masked with the PMA-Ethiopia random numbers in the household dataset. The same random number is applied to the same REs across different surveys of the PMA-Ethiopia grant.

PMA2020_RE_ID: The resident enumerators (REs), or interviewers, who were involved in the previous PMA2020 surveys (2014-2018), also had PMA2020 IDs. This ID is consistent for all survey rounds of the PMA2020 grant.

participant_ID: A unique ID of each respondent participating in panel surveys. The same ID is given to the same female respondent across different surveys. This variable should be used to merge or identify women throughout the study.

SWmetainstanceID: SWmetainstanceID is the unique ID generated by ODK for each female form submitted to the central server. For PMA-Ethiopia, the variable SWmetainstanceID is unique for each female in the six-week follow-up survey.

SWFUweight: Weight for the six-week follow-up survey data

current_methodnum: The numbering scheme for contraceptive methods is consistent across all PMA countries. For example, female sterilization is equal to 1 in every PMA country, whether or not there are any reported uses of female sterilization in the dataset. In some countries, therefore, the numbering will be non-consecutive if some method choices are not selected.

total_live_birth: The number of live births in the pregnancy, if the woman had only still born infants, she will not receive the 6-month or 1-year follow-up.

Combining the data with the Baseline Dataset

For the 5-9 weeks postpartum women (n=273) whose received the 6-week follow-up survey questions at the same time as the baseline interview, their 6-week follow-up data was included in the Cohort 2 Baseline Dataset.

This 6-week follow-up dataset contains only the observations for the following women:

- Women currently pregnant at baseline who consented to the baseline and consented to follow-up (even if they did not complete the baseline)
- Women 0-4 weeks postpartum at baseline who consented to the baseline and consented to follow-up (even if they did not complete the baseline)

Depending on the type of analysis you are doing, you can combine the 6-week follow-up dataset with the Cohort 2 Baseline dataset in different ways.

Generate a Complete 6-week Follow-up Dataset:

To generate a complete 6-week follow-up dataset, you will need to append the women 5-9 weeks postpartum at baseline from the baseline dataset to this 6-week follow-up dataset as follows:

1. Prepare a sub dataset from the Cohort 2 Baseline dataset:
    a. In the baseline dataset, keep only the data of the women 5-9 weeks postpartum at baseline (**baseline_status==3**)
    b. Reference the *PMAET Household & Female Master Codebook* and keep the 6-week follow-up survey variables and the variables of your analytical interest.
    c. In order to appropriate append, you may need to rename some variables.
    d. Save this sub dataset

2. Append to the sub dataset to the 6-week Follow-up dataset and save the combined dataset.

Note: A example Stata dofile for generating the full 6-week follow-up dataset is included for your reference.

Combining Data to Perform Longitudinal Analysis:

If you wish to do longitudinal analysis, you will need to merge data from the baseline to the 6-week dataset. You can do this in two ways:

(1) Merge the full baseline dataset to the 6-week dataset

1. Prepare the baseline dataset
    a. Keep only the data of women whose baseline female survey result is non missing (**FRS_result_cc!=.**).
    b. **participant_ID** variable will be used to merge the dataset. For women who did not complete nor partly complete the survey would have missing value for this variable. Create dummy participant ID for ones that were missing for merging. The dummy participant ID created will be replaced back with missing after merging is complete.
    c. Save the prepared baseline dataset as a temporary file.
2. Prepare the 6-week follow-up dataset
    a. Rename the variables (except for the **participant_ID** merge key variable) by adding prefixes or suffixes to prevent data loss. You may need to rename some variables which had long variable names to meet your statistical package's variable name maximum length requirement.
    b. Save the prepared 6-week follow-up dataset as a temporary file.
3. Merge the two prepared temporary data files – baseline and 6-week follow-up datasets by using the variable **participant_ID**.
4. Replace the dummy participant ID created above (no. 1b) back with missing.
5. Save the merged dataset.

(2) Merge the appended full 6-week dataset with key variables from the baseline dataset.

1. Prepare the baseline dataset:
    a. Keep only the data of women with participant ID (**participant_ID**!="")
    b. Save the prepared baseline dataset as a temporary file.
2. Prepare the full 6-week follow-up dataset
    a. Rename the variables (except for the **participant_ID** merge key variable) by adding prefixes or suffixes to prevent data loss. You may need to rename some variables which had long variable names to meet your statistical package's variable name maximum length requirement.
    b. Save the prepared full 6-week follow-up dataset as a temporary file.
3. Merge the two prepared temporary data files – baseline and full 6-week follow-up datasets by using the variable **participant_ID**. For longitudinal analysis, women whose data was missing for the baseline survey or the 6-week follow-up survey should be dropped from the analysis.
4. Save the merged dataset.

Note: Example Stata dofiles for merging are included for your reference.

GPS Variables

GPS coordinates are not released in this dataset.

Notes for Missing Data

In Stata, Missing data is expressed as "." in the cell. Generally, Stata commands perform computations of any type handle missing data by omitting the row with the missing values. However, this may vary across commands. PMA does not impute missing values. Missing data in datasets should be studied and/or treated before proceeding to analysis.

Reasons for missing data:

Normal situations:

1. Incomplete forms: If a household, female, or SDP form is not marked as completed (HHQ_result, FRS_result, SDP_result not equal to 1), the observation is likely to miss most of the information. Incomplete forms should not be included in the analysis.
2. Observations that are ineligible for subsequent forms: Only eligible respondents will receive subsequent forms. For example, males and ineligible females will not receive female questionnaires in family planning surveys, hence their observations will have all missing values in female forms.
3. Question not administered due to skip logic: PMA surveys use ODK's skip logic function. The subsequent questions are administered selectively based on the respondent's previous answers. Irrelevant or inapplicable questions are skipped. For example, a woman who is not a contraceptive user will not be asked questions about contraceptive usage subsequently.

Uncommon situations:

1. Lost forms: Due to technical constraints in some challenging data collection areas, forms can be lost in the process of data submission. Although most forms were recoverable, there are occasionally a few that cannot be found. For example, an observation from an eligible woman with completed female form information but missing household form information, or vice versa. These observations may be dropped based on analysis needs.
2. Missing due to incorrect skip logic: PMA surveys were conducted under rigorous quality control. However, in rare cases, there can be incorrect skip logic, which skipped a question that was supposed to be administered, resulting in missing values. These errors are documented in the PMA codebook, which can be downloaded from PMA website. It's not necessary to drop the entire observation since this will likely affect only a few questions.

Distinguish missing data from negative values:

1. -99: No response. The respondent was administered with the question but did not provide an answer. PMA survey requires consent from the respondent and the respondent has the right to refuse to answer any questions at any point. -99 is recorded to reflect that the respondent did not provide an answer to a certain question.
2. -88: Did not know. The respondent consented to answer a specific question but without knowing the answer.
3. -77: Not applicable. The question is administered to the respondent but not applicable to the respondent's situation.

PMA GitHub Repository

The PMA GitHub Repository https://github.com/PMA-DM/PMA_Analyses_Public has Stata .do files which could be used to generate indicators in the briefs, using the Household and Female, and Service Delivery Point Datasets that are publicly available.

Note: Data presented in the online briefs represent preliminary results. Therefore, there may be slight differences between the .do file results and those in the brief. Please access the PMA DataLab https://datalab.pmadata.org/ to cross check any discrepancies and get the final estimates.

Dataset Version Updates

Any updates made to datasets after their initial release will be documented here. PMA releases new versions on as needed basis and users who have been approved previously to get access to the datasets will be notified via email upon the release of the new versions. Users can then log onto the PMA website and re-download the datasets without having to submit a new request.

Users should always be using the latest version of the datasets available on the PMA website. However, if users need an access to the old versions of the datasets for any reasons, users can contact datamanagement@pmadata.org directly with an explanation for why the access to the old version was needed.

Dataset Citations

Suggested citation: Addis Ababa University School of Public Health and The Bill & Melinda Gates Institute for Population and Reproductive Health at The Johns Hopkins Bloomberg School of Public Health. Performance Monitoring for Action Ethiopia (PMA-ET) Panel: Cohort 2 - Six-Week Follow-up Survey (Version 1.0), PMAET-Panel-C2-6wkFU. 2022. Ethiopia and Baltimore, Maryland, USA. https://doi.org/10.34976/j5rf-8736.

To report errors or inconsistencies:

Please email datamanagment@pmadata.org.