User Notes for PMA-Ethiopia Panel Cohort 1 (Baseline) Survey Household & Female Dataset, Version 2.0

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.

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 1 Baseline survey used a two-stage cluster design with urban-rural, and major regions as strata. A total of 206 enumeration areas (EAs), located within Panel regions described above, are selected from the master sample frame of the Central Statistical Agency. All pregnant women and post-partum women (age 15 – 49) residing in households, located within the Panel survey regions, were asked to participate in the Panel Survey. A total of 32,614 households completed the census with 32,791 women being screened for the Panel survey. A total 2,868 women (99.3%) completed the Panel survey. Data collection was conducted between September and December 2019.

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

Analytic Sample

PMA-Ethiopia analyses include only observations from completed household interviews. The female sample includes only completed female interviews from completed households. The majority of the indicators includes only de facto women (women who slept in the household the night before). All observations, however, are included in the dataset to allow end users to calculate response rates. For more information on PMA's analytical approach, please refer to PMA Analytical Handbook available at https://www.pmadata.org/data/survey-methodology.

PMA|ETHIOPIA
PERFORMANCE MONITORING FOR ACTION

JOHNS HOPKINS | Bill & Melinda Gates Institute for
BLOOMBERG SCHOOL | Population and Reproductive Health
of PUBLIC HEALTH

Dataset and Questionnaire Structures:

Who is in the data?
There are two types of households in the PMA Ethiopia Baseline data:
- Households identified at census and screening as containing at least one pregnant or 0-9 weeks postpartum woman (panel eligible woman).
- Households selected for the cross-section where during the cross-sectional interview a woman in the household was identified as panel eligible.

Within those households there are four types of women:
- Currently pregnant women who consented to the panel study (**baseline_status**==1)
- Women 0-4 weeks postpartum who consented to the panel study (**baseline_status**==2)
- Women 5-9 weeks postpartum who consented to the panel study (**baseline_status**==3)
- Women identified as eligible at screening but who did not consent to the panel study at the time of baseline or became ineligible for the study between screening and baseline (**baseline_status**==. & **FRS_result**!=.)

Each observation is a household member of the household identified as containing at least one panel eligible woman. All panel eligible women have female forms.

In the female 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 method). 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 **m1**, **m2**, **m3**, **birth1**, **birth2**, **baby1**, and **baby2**, etc...).

Pregnancy History Questions:
Women who were identified as 0-9 weeks postpartum were not asked if they were pregnant and therefore missing for the **pregnant** variable for them.

Contraceptive History Questions:
All panel eligible women were asked about their contraceptive history in the last 2 years with a group of questions repeating for each method usage. The methods were asked in reverse chronological order with variable names starting with **m1_** referring to the most recent method besides the current method.

Postpartum women were asked about their current family planning method use and intentions.

Reproductive Coercion Questions:
All panel eligible women were asked the questions on reproductive coercion and intimate partner violence (IPV).

Antenatal Care (ANC) Questions:
All panel eligible women (**baseline_status**!=.) received the questions on antenatal care (ANC). For questions that were asked differently to pregnant or postpartum women the

suffixes **_preg** and **_pp** are used to differentiate. Questions about plans for delivery or post-delivery were only asked to currently pregnant women (**baseline_status==1**).

Six-week Follow-up Survey:
The first follow-up (6-week Follow-up) survey was to be conducted at 5-9 weeks post-partum and, hence, post-partum women who were enrolled to the panel survey and who at the time of baseline survey were also at 5-9 weeks post-partum (**baseline_status==3**) were eligible for both baseline and 6-week follow-up surveys at the same time. For these (n=277) women, we administered both the baseline questionnaire and 6-week follow-up questionnaires. The variable **cohort_type** (**cohort_type==BL & 6wkFU**) indicates this scenario in the dataset. Consequently, the data for questions about delivery, birth outcomes, neonatal care, and postnatal care (PNC), 6-week follow-up questions, were present only for women 5-9 weeks postpartum (**baseline_status==3**) women.

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.

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 (e.g. if a household respondent said that they use two sources of water, such as a protected well and rainwater, the value of the observation would read "protected_well rainwater"). Multi-select options are generally, though not always, transformed into binary variables for analysis.

Variable Response Options

Select one: Most select one numeric variables have consistent values for option choices across all PMA countries (e.g. marital_status==1 is equivalent to currently married in all PMA countries). Exceptions include the variables **school**, **floor**, **roof**, and **walls**, which have country-specific options and numbering.

PMA|ETHIOPIA
PERFORMANCE MONITORING FOR ACTION

JOHNS HOPKINS | Bill & Melinda Gates Institute for
BLOOMBERG SCHOOL | Population and Reproductive Health
of PUBLIC HEALTH

Select multiple: Similarly, most select multiple variables have the same response options across all PMA countries. Some select multiple variables, however, such as **assets**, have answer options that vary across 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.

svy_design: This variable indicates whether the survey design is cross-sectional or panel (longitudinal).

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.

panel_both: This variable indicates whether the participant was enrolled in panel only surveys or in both cross-sectional and panel surveys.

baseline_status: This variable indicates the participant status at the time of the baseline interview, categorizing them into pregnant, 0-4 weeks postpartum, or 5-9 weeks postpartum. This variable is used to determine whether the participant would receive questions about delivery, birth outcomes, neonatal care, and postnatal care (PNC).

wealth: PMA Ethiopia datasets include **wealthquintile**. The continuous variable **score** is included to allow for construction of various wealth categories.

metainstanceID: metainstanceID is the unique ID generated by ODK for each form submitted to the central server. For PMA-Ethiopia, the variable metainstanceID is unique for each household but will be repeated within the household. **memberID** will provide a unique ID for each person within the household.

FQmetainstanceID: FQmetainstanceID is the unique ID generated by ODK for each female form submitted to the central server. For PMA-Ethiopia, the variable FQmetainstanceID is unique for each female surveyed.

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.

cp, mcp, tcp: Variables that identify current users of any contraceptive method (**cp**), a modern contraceptive method (**mcp**), and a traditional contraceptive method (**tcp**) are included in publicly available datasets so that PMA-Ethiopia estimates involving current contraceptive use and method mix can be replicated. Values for these variables are 0 (no) or 1 (yes). PMA codes **cp**, **mcp**, and **tcp** based on the variable **current_methodnum** with the following caveats:

1. Women who report not being a current user of contraception (**current_user**=0), but who report using EC in the past 12 months are coded as **cp**=1 and **mcp**=1. During analysis, current method is classified as EC in the method mix. The variables **current_methodnum_rc** reflects this.

2. Women who report using LAM as a current method (**current_methodnum**=14. LAM) must satisfy the three conditions listed below to be coded as **mcp**=1. If any of these conditions are not met, these women are coded as **tcp**=1. During analysis, current method is classified as LAM or traditional method. The variable **current_methodnum_rc** reflects this.
    a. Less than six months post-partum
    b. Amenorrheic
    c. Indicating that they are using LAM with the intention of preventing pregnancy

GPS Variables

GPS coordinates are not released in this dataset.

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.

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

PMA|ETHIOPIA
PERFORMANCE MONITORING FOR ACTION

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Bill & Melinda Gates Institute for
Population and Reproductive Health

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.

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.

In March 2021, Version 2 of this dataset was released with the following updates:
- **participant_ID**, **any_method_2yr**, **method_in2yrs** variables were added
- **birth_urban**, **prior_urban**, and **ur** variable coding and their associated **urban_rural_list** ListOfChoice were corrected.
- **strata** variable coding was correct.

- As part of the standardization effort to make variable naming consistent across PMAET, PMA Core and PMA2020 surveys, the naming of some of the variables and ListOfChoice were changed. The codebook that is compatible with this updated dataset can be re-downloaded from https://www.pmadata.org/data-codebooks.

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) Household and Female Panel: Cohort 1 - Baseline Survey (Version 2.0), PMAET-Panel-C1-Baseline.* 2019. Ethiopia and Baltimore, Maryland, USA. https://doi.org/10.34976/h75w-8084.

To report errors or inconsistencies:

Please email datamanagment@pmadata.org.