User Notes for the PMA India (Rajasthan) Phase 3 Household and Female 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.

PMA Survey Design

Performance Monitoring for Action (PMA), formerly PMA2020, builds on the previous success of PMA2020 surveys in India (Rajasthan) and focused on collecting routine data on key global indicators in family planning and reproductive health, while expanding content area to address questions of contraceptive decision-making and autonomy, in order to better understand the determinants and consequences of unique contraceptive use and patterns of use in India (Rajasthan). These are measured through three-related data collection activities: household and female surveys (HQFQ) producing both cross-sectional and longitudinal data, Service Delivery Point panel surveys (SQ), and an SDP client exit surveys (CQ).

In India (Rajasthan), a cross-sectional and panel household and female surveys (HQFQ) are conducted annually, with follow-up for the panel occurring at Year 2 and 3. The Service Delivery Point Survey (SQ) panel baseline data is collected at Year 1 and follow-up data will be collected annually. The Service Delivery Point Client Exit Survey (CQ) is conducted biannually with a baseline and a phone interview follow-up occurring 6 months after the baseline enrollment each year.

PMA survey uses a multi-stage cluster design, with stratification at the urban and rural level and/or by sub-region. The enumeration area (EA) is the primary sampling unit, obtained from the national statistics agency of the respective geography. Within each urban/rural or sub regional stratum, EAs are selected using probability proportional to size (PPS) method. In each of the EAs, all households and private health facilities are listed and mapped prior to baseline data collection. Listings of public health facilities that serve the selected EAs at all three levels are obtained from the Ministry of Health.

For household and female cross-sectional and panel surveys, resident enumerators (RE) annually conduct a full listing of households within each enumeration area (EA). The annual listing is used to update the baseline weights to generate the cross-sectional estimates. At baseline, 35 households were randomly selected within each EA for interview. RE administered a household questionnaire, including completing a census of household members and guests who slept there the previous night for all selected households who consented to participate. Eligible female aged 15-49, who were usual members, were consented to participate in the female panel survey. Guests who slept at the households were included in the female cross-sectional survey.

PMA uses an open panel design, enrolling new eligible women at annual follow up (year 2 and year 3). Households selected at baseline and still residing in the study area will be followed-up in subsequent rounds. The study area for surveys was accessible parts of the entire county, and other PMA counties if women moved outside of her Phase 1 county. Adolescents in selected households who were 14 years in the Adolescents in selected households who were 14 years in the previous round will be enrolled in the panel as 15-year-olds starting in Year 2. Women who were 49 years at an earlier round will not be interviewed in subsequent rounds. Households who moved out of the study area since baseline will be considered lost-to-follow-up. New households residing in residential structures of households interviewed at baseline will be enumerated and enrolled in the panel in subsequent rounds. In addition,

when an initially sampled housing structure is vacant or demolished before Year 2 or Year 3 rounds, a new housing structure (dwelling unit) will be randomly selected form the new household listing to replace the lost one.

Sampling

PMA sampling approach is structure-based that is, PMA considers each physical structured HH (or dwelling) that was randomly selected at Phase 2 or Phase 1 as the fixed panel. PMA returned to the same EAs and dwelling units that were listed during the Phase 2 or Phase 1 HQFQ survey.

Who is in the Data?

India (Rajasthan) Phase 3 (UGP3) Household and Female Survey collected data from households living in three different types of dwellings to capture cross-sectional and panel samples, and the information collected for each type of dwelling unit is slightly different. **HH_type** variable identifies these four types of dwellings.

Household Survey:

1. **Phase 2 Dwellings** (**HH_type**==1)**:** these are the same physical dwellings where the households that PMA interviewed at Phase 2 lived. These are part of the cross-sectional sample. Within a Phase 2 dwelling, there are two types of households.
   a. Phase 2 Household: If at least one member from the household that lived in the dwelling at Phase 2 is still a resident, PMA considers the household to be the same household as Phase 2. In this case there are three possible types of household members.
      i. Resident Phase 2 Member: A member of the household from Phase 2 who is still a resident. The resident enumerator (RE), interviewer, has the chance to update their age, relationship to household head, and marital status.
      ii. Non-Resident Phase 2 Member: A member of the household from Phase 2 who is no longer a resident. The RE updates their observation with information on why they no longer live with the household. The only exception to this is if the member is a panel woman found in another dwelling, in that case her observation from her Phase 2 household is dropped to avoid duplicates.
      iii. New Household Members: Someone who has become a member of the household between Phase 2 and Phase 3. They are added to the household roster using PMA's standard household roster questions.
   b. New Household: If none of the members from the household that lived in the dwelling at Phase 2 still live in the dwelling, PMA considers the dwelling as occupied by a new household. The RE then answers questions on what happened to the Phase 2 household before proceeding through PMA's standard household questionnaire.
2. **Replacement Cross-Section Dwellings** (**HH_type**==2): In EAs (or Clusters) where >10% of the Phase 2 Dwellings were lost to follow-up (LTFU), PMA selected new dwellings to be interviewed equal to the number of dwellings LTFU. These are part of the cross-sectional sample. In these cases, the RE completes the standard PMA household questionnaire.
3. **Panel Women Follow-Up Dwellings** (**HH_type**==3): A dwelling within the study area where a panel woman who moved out of her Phase 1 dwelling currently lives (and that dwelling was not another Phase 1 dwelling nor selected to replace a LTFU dwelling). These are NOT part of the cross-sectional sample. Within these dwellings there are two types of household members.

PMA
PERFORMANCE MONITORING FOR ACTION

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

   a. <u>Panel women</u>: The panel women from Phase 1 who live in the dwelling. The RE has the chance to update their age, relationship to household head, and marital status.
   b. <u>New Household Members</u>: Other members of the household currently residing in this dwelling between Phase 2 and Phase 3. They are added to the household roster using PMA's standard household roster questions.

Female Survey:

1. **Cross-Section Only Women:** These women received only the baseline female questionnaire and cannot be linked back to the data at Phase 1 or Phase 2. These women are included in cross-sectional analysis only.

   There are three scenarios when a woman would be part of this sample.
   a. <u>Aged in from Phase 1</u>: A Phase 2 household member who was 14 at Phase 2 and has now aged into the eligibility criteria.
   b. <u>Refused follow-up at Phase 2</u>: If a woman refused follow-up at Phase 2 FQ and was 48 or younger at Phase 2, she was given the opportunity to consent to the Phase 3 baseline.
   c. <u>New Woman 15-49</u>: Any woman 15-49 found in a cross-section dwelling who was not interviewed at Phase 2.

2. **Cross-Section and Panel Women**: Women 15-48 from Phase 2 who consented to follow-up and were found at Phase 3 in a cross-section dwelling. These women are included in both cross-sectional and panel analysis.

3. **Panel Only Women**: Women 15-48 from Phase 2 who consented to follow-up and were found at Phase 3 in a dwelling not selected for the cross-section. These women are included in panel analysis only.

Phase 3 Cross-sectional Sample

PMA India (Rajasthan) Phase 3 Household and Female Survey collects data in 134 enumeration areas (EAs) using a multi-stage stratified cluster design with urban-rural strata. The results are representative at the national level and within urban/rural strata. The final sample included 4,364 households and 5,481 *de facto* females who completed the interview for the cross-sectional sample. Data collection was conducted between September and October 2022.

Phase 3 Female Panel Sample

HH Loss to Follow-up (LTFU): A Phase 2 or Phase 1 dwelling unit was considered LTFU if the household questionnaire result was "Vacant/Not a dwelling unit", "Dwelling destroyed", and "Dwelling not found".

Panel Woman Loss to Follow-up (LTFU): Panel women LTFU can take place at one of four different "levels" or timepoints: the dwelling unit level, the household (HH) level, the household roster level, or the female questionnaire (FQ) level.

Dwelling unit level LTFU happened when the women's entire household moved out of the Phase 2 or Phase 1 dwelling unit, with some households moving within the study area, some others moving outside of the study area, or because PMA was unable to acquire any information on the household from key informants, such as neighbors or village/town heads.

PMA
PERFORMANCE MONITORING FOR ACTION

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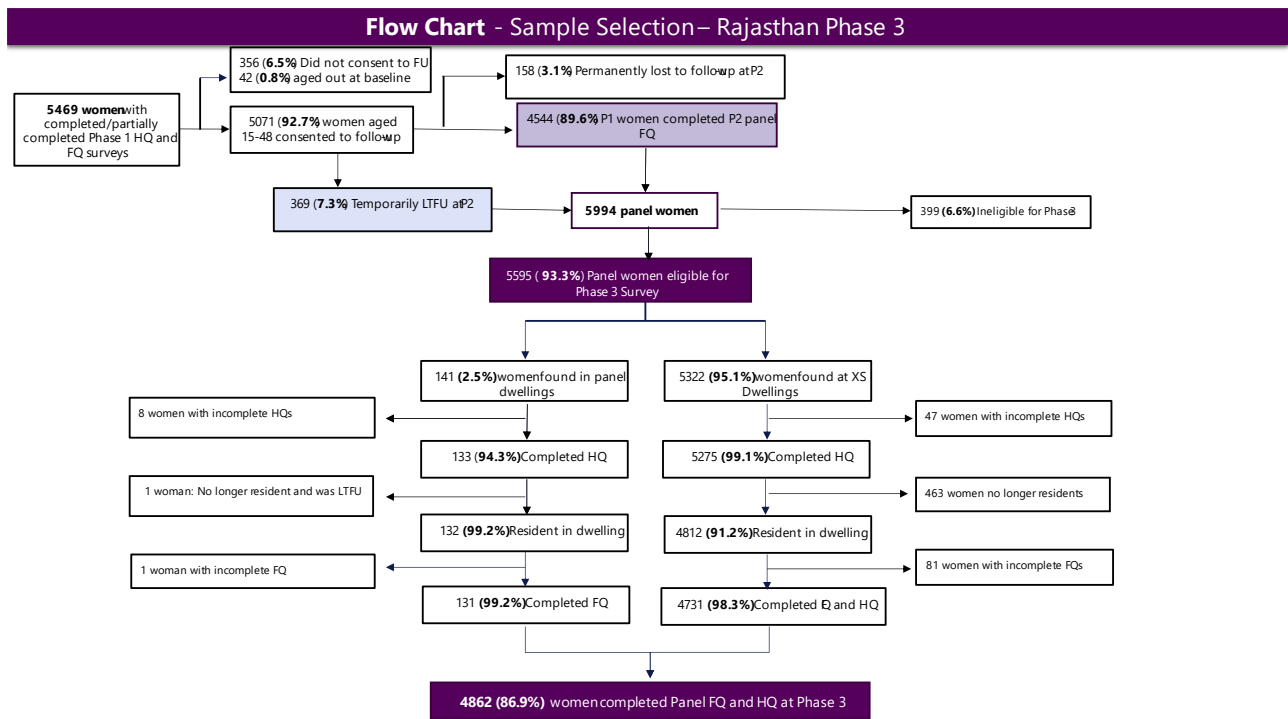Population and Reproductive Health

Household (HH) level LTFU happened when there was an incomplete household survey (HQ) due to refusal to participate, incomplete interviews or unavailability/absence of a household respondent during the study period.

Household roster level LTFU happened when the panel women were no longer residents based on the roster status. They were either found to have died since Phase 2 or Phase 1 surveys, to have moved out of the study area for school or other reasons, or whose status was simply unknown.

Female questionnaire (FQ) level LTFU happened when panel women did not complete the phase 3 female interview.

Panel Sample:

Phase 3 panel data collection was conducted among 5,595 females aged 15-49 at the time of the Phase 3 survey who were interviewed at Phase 1 or Phase 2 and consented to follow-up. Among the Phase 3 panel women, 83.1% of them were enrolled in the panel survey at Phase 1 between August and October 2020 and 16.9% were enrolled in the panel survey at Phase 2 between September and December 2021. Of the 5,595 eligible respondents, 11.8% were not reached for follow-up at Phase 3. Of those reached, 4,862 (86.9%) completed the Phase 3 survey, for a response rate of 98.3% among contacted women.

Flow Chart - Sample Selection – Rajasthan Phase 3

5469 women with completed/partially completed Phase 1 HQ and FQ surveys	

356 (**6.5%**) Did not consent to FU
42 (**0.8%**) aged out at baseline

5071 (**92.7%**) women aged 15-48 consented to follow-up

158 (**3.1%**) Permanently lost to follow-up at P2

4544 (**89.6%**) P1 women completed P2 panel FQ

369 (**7.3%**) Temporarily LTFU at P2

5994 panel women

399 (**6.6%**) Ineligible for Phase 3

5595 (**93.3%**) Panel women eligible for Phase 3 Survey

141 (**2.5%**) women found in panel dwellings

5322 (**95.1%**) women found at XS Dwellings

8 women with incomplete HQs

133 (**94.3%**) Completed HQ

5275 (**99.1%**) Completed HQ

47 women with incomplete HQs

1 woman: No longer resident and was LTFU

132 (**99.2%**) Resident in dwelling

4812 (**91.2%**) Resident in dwelling

463 women no longer residents

1 woman with incomplete FQ

131 (**99.2%**) Completed FQ

4731 (**98.3%**) Completed FQ and HQ

81 women with incomplete FQs

4862 (86.9%) women completed Panel FQ and HQ at Phase 3

Imputations and Replacements

Since we had relevant data for panel women from the Phase 2 survey, we did not ask them questions that would not have changed in the last year, such as birth date, date of first birth, date of first marriage etc. To make sure we had all the data for PMA's standard analyses, we replaced those missing values in the Phase 3 dataset with the values from Phase 2.

Materials Included in the Data Zipped File

This data zipped file includes:
1. Household questionnaire pdf file
2. Female questionnaire pdf file
3. User notes pdf file
4. Dataset in 3 file formats: .csv, .xlsx and Stata .dta

Codebook

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

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**).

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.

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 PMA Household and Female Survey Master Codebook for complete details on variables and answer choices for each survey.

Specific Variables

metainstanceID: A unique ID generated by ODK for each household survey form submitted to the central server. It 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: A unique ID generated by ODK for each female survey form submitted to the central server. Only female participants who were selected for the female survey had values for

FQmetainstanceID and it is unique for each female within a survey but different across the survey phases. FQmetainstanceID of the Phase 1 Survey is used to generate **female_ID**, which is a panel merge key variable.

female_ID: This is a unique ID of the female participants across different phases of the PMA survey. This variable, hence, serves a panel merge key variable and can be used to combine different phases of the datasets for panel analysis. This variable is generated from the FQmetainstanceID of the Phase 1 Household and Female Survey.

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

RE_ID: Identification number of the resident enumerator (RE), or interviewer. RE names are masked with unique numbers. The same unique number is applied to the same RE across different phases of the PMA survey.

PMA2020_RE_ID: For the resident enumerators (REs), or interviewers, who were involved in the previous PMA2020 surveys (2014-2018), this variable provides their IDs from PMA2020 surveys. This ID is unique across different surveys.

wealth: The dataset includes **wealthquintile** and/or **wealthtertile** depending on the country. The continuous variable **score** is included to allow for construction of various wealth categories.

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 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 (8. Emergency contraception) in the past 12 months in the *Contraceptive Calendar* 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**.
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

Contraceptive Calendar

The Contraceptive Calendar period for India (Rajasthan) Phase 3 survey was 36 months (2020-2022). The Contraceptive Calendar consists of two columns. Column 1 records births, pregnancies, terminations and contraceptive use, while column 2 records reasons for discontinuation of contraceptive use.

calendar_c1_full, **calendar_c2_full**: These are the contraceptive calendar column 1 and column 2 data prepared as string variables in a reverse chronological order from the left end. Any months that occur in the future in reference to the time of the interview during the calendar period has no data and are denoted with just commas. The codes for the options of column 1 and column 2 are listed in the table below.

	Column 1		Column 2
Code	**Births, Pregnancies or Contraceptive Use**	**Code**	**Reasons for Discontinuation**
B	Births		
P	Pregnancies		
T	Terminations		
0	No method used	1	Infrequent sex / husband away
1	Female Sterilization	2	Became pregnant while using
2	Male Sterilization	3	Wanted to become pregnant
3	Implant	4	Husband / partner disapproved
4	IUD	5	Wanted more effective method
5	Injectables	6	Side effects / health concerns
7	Pill	7	Lack of access / too far
8	Emergency Contraception	8	Costs too much
9	Male Condom	9	Inconvenient to use
10	Female Condom	10	Up to god / fatalistic
11	Diaphragm	11	Difficult to get pregnant / menopausal
12	Foam / Jelly	12	Marital dissolution / separation
13	Std Days / Cycle Beads	96	Other
14	LAM		
30	Rhythm method		
31	Withdrawal		
39	Other traditional methods		

Panel Key Variables

Variable name	Definition
HH_type	Categorical variable that flags the type of dwelling.

pr_dwellingID	The metainstanceID associated with the *dwelling* the first time an HQ was submitted for the dwelling. A P1 dwelling interviewed at Phase 2 (P2) and Phase 3 (P3) will have the same pr_dwellingID and it will be the metainstanceID from Phase 1 for the dwelling. A new P2 dwelling interviewed at P2 and P3 will have the same pr_dwellingID and it will be the metainstanceID from Phase 2 for the dwelling.
pr_metainstanceID	The metainstanceID associated with the *household member* the first time the member was interviewed. It links the people to their baseline dwelling. If someone lives in the same dwelling at Phase 3 as their baseline dwelling, pr_metainstanceID would be the same as pr_dwellingID. If the person has moved to a different dwelling, pr_metainstanceID could be used to merge the person to their baseline dwelling in the Phase 1 or Phase 2 dataset.
pr_hh_status	The status of the Phase 2 household if they no longer live in the Phase 2 dwelling.
hh_member_status	The status of the Phase 2 household member in the Phase 2 dwelling, at Phase 2.
household_replaced	This variable flags Phase 2 cross-sectional dwellings that were LTFU and replaced in EAs with a household LTFU rate > 10%.
pr_fq_id	The baseline FQmetainstanceID(s) of the panel women selected for interview in a panel woman follow-up dwelling
female_ID	The baseline FQmetainstanceID of the panel woman. This variable should be used to merge panel women across the Phase 1, Phase 2, and Phase 3 datasets. A panel woman enrolled at P1 and interviewed at P2 and P3 will have the same pr_FQmetainstanceID and it will be the FQmetainstanceID from Phase 1 for the woman. A panel woman enrolled at P2 and interviewed at P3 will have the same pr_FQmetainstanceID and it will be the FQmetainstanceID from Phase 2 for the woman.
origPRpanel_dwelling	Among dwellings sampled at Phase 2 and returned to at Phase 3, a binary variable flagging if any panel women lived in the dwelling at Phase 2.

PMA
PERFORMANCE MONITORING FOR ACTION

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

cur_panel_dwelling	Among all dwellings sampled at Phase 3, a binary variable flagging if any panel women live in the dwelling at Phase 3. Use this variable to get a dataset of all households and household members living in a dwelling with a panel woman at Phase 3.
xs_sample	A binary variable flagging if an observation is part of the cross-section. To restrict your analyses to the cross-sectional sample, only keep the observation with xs_sample=1 (Eg Stata code: `keep if xs_sample==1`).
panel_woman	A binary variable flagging if the observation is a panel woman. To use panel women only in your analyses only keep the observation with panel_woman=1 (Eg Stata code: `keep if panel_woman==1`). This does not include newly added women consented to phase 3.
newly_enrolled	A binary variable flagging if the cross-section only woman consented to future follow-up.
HHweight	The household weight for cross-section households.
FQweight	The female weight for cross-sectional analysis.
P1P3_FUweight	The female weight for panel analysis of women who completed P1 and P3.
P2P3_FUweight	The female weight for panel analysis of women who completed P2 and P3.
P1P2P3_FUweight	The female weight for panel analysis of women who completed P1, P2, and P3.
wealth	The wealth tertile/quintile of the dwellings in the cross-sectional sample only. Panel only dwellings will not have a value for wealth. (Panel only dwellings happen when they move out of dwelling unit outside of EA).

Combining the Data for Longitudinal Analysis

If you wish to perform longitudinal analysis with the India (Rajasthan) Phase 2 and Phase 1 Household and Female Survey datasets, you can combine the datasets using the merge key variable **female_ID**. Before merging the datasets, restrict the sample to panel women (**panel_woman**==1), drop the cross-sectional only women, and add prefixes or suffixes to your variables of interest to prevent data loss.

Phase 3 Data Imputations

Since we had relevant baseline data for panel women from the Phase 2 survey, we did not ask them questions that would not have changed in the last year, such as birth date, date of first birth, date of

first marriage etc. However, to make sure we had all the data for PMA's standard analyses, we replaced those missing values in the Phase 3 dataset with the values from Phase 2.

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** and **CEI_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.

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.

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, Service Delivery Point, and Client Exit Interview 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 Citation

Suggested citation: Indian Institute of Health Management Research (IIHMR) University in Jaipur; the Bill & Melinda Gates Institute for Population and Reproductive Health at the Johns Hopkins Bloomberg School of Public Health; and Jhpiego. Performance Monitoring for Action (PMA) India (Rajasthan) Phase 3: Household and Female Survey (Version #), PMA2022/IN-P3-HQFQ. 2022. India and Baltimore, Maryland, USA. https://doi.org/10.34976/zwsv-md35

To report errors or inconsistencies:

Please email datamanagement@pmadata.org