

User Notes for PMA India (Rajasthan) Household and Female Phase 1 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) is conducted biannually with a baseline and a follow-up occurring 6 months after the baseline enrollment each year. The Service Delivery Point Client Exit Interview Survey (CQ) baseline data is collected at Year 1 and follow-up data will be collected annually.

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 EA will be followed-up in subsequent rounds. 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 EA 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 household will be randomly selected from the new household listing to replace the lost one.

Sampling

India (Rajasthan) Phase 1 (IN-RJP1) Household and Female Baseline Survey includes 134 enumeration areas (EAs) selected using a multi-stage stratified cluster design with urban-rural strata. The results representative at the state level and within urban/rural strata. The final sample included 4,577 households and 5,405 females which completed the interview. Data collection was conducted between August and November 2020.

Analytic Sample

Analyses include only observations from completed household interviews. The female sample includes only completed female interviews from completed households. The majority of indicators include 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.

Materials included in this 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>.

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**).

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

Specific Variables

metainstanceID: metainstanceID is the unique ID generated by ODK for each 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: FQmetainstanceID is the unique ID generated by ODK for each female 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 and different across the survey phases. FQmetainstanceID of the Phase 1 Survey is used to generate **female_ID**, which is a panel merge key variable.

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 phases.

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

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.

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 1 survey was 36 months (2018-2020). 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	Code Reasons for Discontinuation

Births, Pregnancies or Contraceptive Use

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 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 Survey.

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.

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: 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) Household and Female Survey Phase 1, PMA2020/India-P1-HQFQ*. 2020. India and Baltimore, Maryland, USA. <https://doi.org/10.34976/yk3f-nz21>

To report errors or inconsistencies:

Please email datamanagement@pmadata.org