

User Notes for the PMA Kenya Phase 2 Service Delivery Point Baseline 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 Kenya 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 Kenya. 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 Kenya, 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 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 the service delivery point (SDP) panel survey, up to three private SDPs are selected within each EA at baseline. Public SDPs at the three levels (primary, secondary, and tertiary) of the healthcare system assigned to serve the sampled EA are also included in the sample. This baseline sample of SDPs is followed-up annually at year 2 and 3. Private SDP interviews are conducted by REs, public SDP interviews are conducted by supervisors with facility in-charges, who provide consent for their facility to participate. Other personnel may participate in the survey as well, depending on in-charge knowledge and facility size.

The SDP panel surveys provide data to monitor service availability and quality, understand attributes for the availability and readiness, and explore the association of these SDP results with individual behaviors in the EA, by geospatially linking data from both the SDP and household surveys. The SDP panel surveys linked to the household panel surveys will inform government policies and programs regarding the influence of service environment on service utilization behaviors among a population.

Sampling

Kenya Phase 2 (KEP2) Service Delivery Point Survey includes 308 enumeration areas (EAs) of selected counties. The EAs were drawn using the same stratified cluster design with urban-rural strata by the Kenya Bureau of Statistics following the pre-existing sampling frame used for PMA2020. The results are national and county-level representative. The final sample included 924 SDPs which completed the interview. Data collection was conducted between November and December 2020.

Materials included in this Data Zipped File

This data zipped file includes:

1. Service delivery point questionnaire pdf file
2. User notes pdf file
3. Dataset in 3 file formats: .csv, .xlsx and Stata .dta

Codebook

The latest version of the PMA Service Delivery Point 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. **today** and **todaySIF**).

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 service delivery point respondent said that the community health volunteers provide male condoms, pill and injectables, the response for this variable would read “male_condoms pill injectables”). 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 which have country-specific options and numbering. For examples:

- Geographic variable (e.g. **region, county**): geographic variable names and response options vary across countries
- **facility_type**: facility types vary across countries

Select multiple: Similarly, most select multiple variables have the same response options across all PMA countries, with exceptions for variables with choices that are country specific.

See the *PMA Service Delivery Point Master Codebook* for complete details on variables and answer choices for each survey.

Specific Variables

metainstanceID: A unique ID generated by ODK for each SQ survey form submitted to the central server. This variable is unique for each facility within a survey.

facility_ID: Facility names are masked with unique ID numbers. This variable is unique across different PMA surveys. Hence, facility_ID is also merge key for combing the data with other datasets for panel analyses.

PMA2020_facility_ID: For facilities that were included in the previous PMA2020 (2014-2018) surveys, this variable provides their PMA2020 IDs. Th observation would be missing for new facilities which were not part of the PMA2020 surveys. For ones that have PMA2020 IDs, it is unique across different surveys.

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 PMA survey 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 PMA survey phases.

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.

EAserved#: Some SDPs serve more than one EA. The EAserved# variables indicate the additional EAs that a given facility serves, if any. Information regarding which EAs an SDP serves comes from the country/local government. Only public facilities are assigned to serve more than one EA.

Combining the Data with Other Datasets

The merge key variable **facility_ID** can be used to combine the Phase 2 SQ data with Phase 1 SQ Baseline dataset. Of the 993 SDPs in the Phase 2 dataset, 960 merges with the Phase 1 dataset. The 33 SDPs not merging to the Phase 1 dataset are the new SDPs added to this phase.

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 Citations

Suggested citation: International Centre for Reproductive Health Kenya (ICRHK) and The Bill & Melinda Gates Institute for Population and Reproductive Health at The Johns Hopkins Bloomberg School of Public Health. *Performance Monitoring for Action (PMA) Kenya Phase 2: Service Delivery Point Baseline Survey (Version 1.0), PMA/Kenya-P2-SQ-Baseline*. 2020. Kenya and Baltimore, Maryland, USA.
<https://doi.org/10.34976/jzd9-9a35>

To report errors or inconsistencies:

Please email datamanagement@pmadata.org