User Notes for PMA Democratic Republic of Congo (Kinshasa & Kongo Central) Phase 1 Household and Female Survey Dataset, Version 1

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

Performance Monitoring for Action (PMA), formerly PMA2020, builds on the previous success of PMA2020 surveys in Democratic Republic of Congo 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 Democratic Republic of Congo. 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 Democratic Republic of Congo, 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) and Client Exit Survey (CQ) are conducted biannually with a baseline and a follow-up occurring 6 months after the baseline enrollment each year.

Sampling

Democratic Republic of Congo Phase 1 (CDP1) Household and Female Baseline Survey includes 58 enumeration areas (EAs) in Kinshasa and 52 EAs in Kongo Central. The EAs were drawn using the stratified cluster design from the master sampling frame using probabilities proportional to size. The results are representative at province level. The final sample included 1,968 (97.0%) households and 2,611 (95.3%) de facto women in Kinshasa, and 1,965 (98.5%) households and 1,950 (98.8%) de facto women in Kongo Central, which completed the interviews. Data collection was conducted between December 2019 and January 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.

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 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.

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_recent_methodnum, current_methodnum, recent_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). PMA2020 and 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 (**recent_methodnum**=8. emergency) 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** and **recent_methodnum_rc** reflect 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

3. Women who report female sterilization as their first contraceptive method (**first_methodnum**=1. female sterilization), but who do not report currently using female sterilization are coded as **cp**=1 and **mcp**=1. During analysis, current method is classified as female sterilization in the method mix. The variable **current_methodnum_rc** reflects this.

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.

Dataset Citations

Suggested citation: Tulane University School of Public Health, University of Kinshasa 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 and Accountability (PMA) Household and Female Phase 1 Baseline Survey, PMA/DRC-HQFQ-P1-BL (*Kinshasa & Kongo Central). 2020. Kinshasa, DRC and Baltimore, Maryland, USA. doi:10.34976/x3bk-0w75.

To report errors or inconsistencies:

Please email datamanagement@pma2020.org