User notes for PMA2016/Burkina Round 4 Household and Female data, version 4

Disclaimer: PMA2020 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.

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**). For all questions requiring a date entry, if the respondent answered either "Do Not Know" or refused to answer the question, the date was recorded as January 1, 2020.

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 countries (e.g. marital_status==1 is equivalent to currently married in all countries). Exceptions include the variables **school**, **fp_provider**, **floor**, **roof**, and **walls**, which have country-specific options and numbering.

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

See the HHQFQ Master Codebook for complete details on variables and answer choices for each PMA2020 country and survey round.

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 rounds.

wealth: Some country datasets include **wealthquintile** while others include **wealthtertile**. 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 PMA2020, 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 PMA2020, 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 PMA2020 countries. For example, female sterilization is equal to 1 in every PMA2020 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 PMA2020 estimates involving current contraceptive use and method mix can be replicated. Values for these variables are 0 (no) or 1 (yes). PMA2020 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. For some PMA2020 countries a separate dataset of displaced GPS coordinates at the EA level is released.

Sampling

PMA2016/Burkina Round 4 used a two-stage cluster design with urban and rural strata. A sample of 83 enumeration areas (EAs) was drawn from the INSD master sampling frame. Each EA was listed and mapped; 35 households were randomly selected. Occupants in selected households were enumerated and eligible females of reproductive age (15-49) were contacted and consented for interview. Data collection was conducted between November, 2016 and January, 2017. The final sample size is 2,751 households and 3,252 females.

Analytic sample

PMA2020 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.

Dataset version updates

Any updates made to datasets after their initial release will be documented here.

In January 2017, all previously released datasets were modified as below:

1. The value of **age_at_first_use_children** is 0 for women who have ever used family planning and who have never given birth. Previously, such women had a missing value for age_at_first_use_children.
2. The values for **water_sources_main_drinking** and **water_sources_main_other** equal the value of water_sources_all if a household has one water source. Previously, such households may have had a missing value for these variables.
3. The value for **sanitation_main** equals the value of sanitation_all if a household has one sanitation facility. Previously, such households may have had a missing value for this variable.

All datasets released after January 2017 will have these changes included.

In July 2017 version 2 of this dataset was released. Modifications from version 1 are as follows:

- The variables **long_last_singlerod, dissolving_implant, IUD_hormone, permanent,** and **IUD_nohormone** were modified so incorrect response values numbered 5 were recoded to -99.

In October 2017 version 3 of this dataset was released. Modifications from version 2 are as follows:

- The variable **total_numbers** was added to the dataset.

In November 2018, version 4 of this dataset was released. Updates from the previous version are as follows:

- The variables **main_drinking_rc** and **sanitation_main_rc** were dropped from the dataset.
- Date variables ending in **SIF** were changed to missing if the corresponding non-SIF values were 01Jan2020.
- Values for the contraceptive method current use variables (**femalester** – **othertrad**) were changed from missing to zero or one if the corresponding value of the variable **cp** was not missing.
- The value of **EA_ID** was replaced with 7054 if it was 7374 to be consistent with the EA identification number in previous rounds.
- The variable implant_removed_attempt_12 was renamed to **implant_removed_attempt** to be consistent with PMA2020 data in other countries.

To report errors or inconsistencies:

Please email datamanagement@pma2020.org