PMA2020 General Sampling Strategy

General sampling procedures for PMA2020

The sample size requirements for PMA2020 are determined by the sample size needed to calculate the modern contraceptive prevalence rate (mCPR) for all women at a given level of precision (margin-of-error). The most recently measured mCPR, generally from the DHS, is used as a benchmark to determine the necessary sample size at the national level. In some cases, the sample size is increased to allow for sub-national estimates. Each country-specific sampling memo details how each sample size was calculated, the level of precision, and the regions at which the sample is representative.

As simple random sampling (SRS) is not preferred for conducting a national level survey due to budget and logistical constraints, we employ multistage stratified clustered sampling, where households are surveyed in clusters or enumeration areas (EAs). All women of reproductive age 15-49 years are targeted for interview. The formula used to determine the final sample size is

$$n = \frac{z^2_{1-\frac{\alpha}{2}}\, p(1-p)}{\delta^2 * R_i * R_j} X\ DEFF$$

where n is the sample size of women;
Z is the abscissa of normal curve (at α=0.05, Z=1.96)
DEFF is the design effect due to multi-stage stratified cluster sampling (a maximum of 3.0 is imposed);
P is the estimated (expected) mCPR;
δ is the desired margin of error;
R_i is the individual response rate;
R_h is the household response rate

Boundaries of selected EAs are mapped and all occupied households within the selected EA are listed. Once all households within the EA boundary are listed, a fixed number of households (HH) within the EA are randomly selected and interviewed. The fixed number of households per EA is typically 35, ranging from 33 to 44. All women age 15-49 who are either usual members (*de jure* population) of the household or who slept in the household the night before (*de facto* population) are included in the women sample. Therefore, no adjustment is made for the probability of selecting an individual from within the household. All analyses based on the female questionnaire are based on *de facto* population.

PMA2020 Survey Sampling Assumptions

1. The mCPR estimate among all women from the most recent DHS has remained constant when the PMA survey was undertaken. mCPR is calculated separately for both national, urban/rural, and when appropriate, the sub-regional level.
2. PMA has same DEFF with the most recent DHS. When less than 3.0, the exact DEFF is used. If greater than 3.0, DEFF is capped at 3.0 in the calculations for keeping the sample size reasonable and logistically feasible to enroll. DEFF is calculated separately for both national, urban/rural, and when appropriate, at the sub-regional level.
3. The response rate is assumed to be cumulatively equal (across both household and female) to approximately 90% (i.e. 95% response rate for household interviews and 95% response rate for female interviews). It is applied consistently across all strata.
4. The average number of women per HH from the most recent DHS survey has remained constant. When possible, it is calculated separately for national, urban/rural and when appropriate, at the sub-regional level. Otherwise, the national estimate is applied uniformly across strata.

If alternate assumptions (such as strata specific women per HH or response rates) can be justified, it is fine to use those in sample size calculation. If alternate assumptions are used and result in a larger number of EAs needed to obtain the stated margin of error, the more conservative (i.e larger) number of EAs should be used, budget permitting.

PMA2020 Sample Selection

When drawing the initial sample, PMA requests that the national sampling organization/agency, generally the National Statistical Office or equivalent, draw the index EA sample using probability proportional to size (PPS) within the designated strata. The statistical office should provide a spreadsheet that includes the list of index EAs and their geographic information, their urban/rural designation, final EA selection probability, and all relevant information used to calculate the EA selection probability (strata, population/number of HHs in strata, number of EAs sampled per strata, county or regional selection probability if relevant).

Additionally, we ask that at the same time the index EAs are drawn, the statistical office identify all geographically contiguous EAs that share the same urban/rural designation as the original index EA. The same information (EA selection probability, strata, population/number of HHs in strata, number of EAs sampled per strata, county or regional selection probability if relevant) should be provided for the geographically contiguous EAs as the original index EAs.

It is critical to clearly identify which EA is the index EA and which contiguous EAs are linked to which index EA in case the EAs are combined for constructed a larger primary sampling unit with a desired number of HHs.