PMA2020 Household and Female Survey Sampling Strategy in Nigeria

The first section describes the overall survey design and sample size calculation method of the Performance, Monitoring and Accountability 2020 (PMA2020) household and female surveys. Later sections detail the parameters of the sample size calculations and the states for which the sample is representative specific to the country. The last section provides methods regarding post-stratification weights to calculate national-level estimates, unique for Nigeria PMA2020.

1. General sampling procedures for PMA2020

The sample size requirements for PMA2020 are determined to calculate the modern contraceptive prevalence rate (mCPR) for all women at a desired level of precision (margin of error). The most recently measured mCPR, generally from the Demographic Health Survey (DHS), is used as a benchmark to determine the necessary sample size at the national level. Usually the sample size is increased to allow for sub-national estimates.

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), which are selected first—with a probability proportional to size (PPS) method--and independently within sampling strata. All women of reproductive age 15-49 years are the targeted population for interviews. The formula used to determine the final sample size of women is:

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

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; and

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 (constant take size) of households (HH) within the EA is randomly selected for interviews. The fixed number of households per EA are typically 35, but can range across PMA survey sites from 33 to 44. All women aged 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 female sample. No adjustment is made for the probability of selecting an individual woman from within the household. Of note is that all PMA2020 survey results based on the female questionnaire are based on analyses of *de facto* population.

PMA2020 Survey Sampling Assumptions

The following assumptions are made when calculating the PMA2020 survey sample size:

1. The mCPR estimate among all women has not changed since the most recent DHS to the time the PMA survey is undertaken. The mCPR for all women 15 to 49 is calculated separately for both national, urban/rural, and when appropriate, the sub-regional level.
2. PMA has same DEFF as 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 to keep the sample size reasonable and logistically feasible to implement. 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) at approximately 90% (i.e. 95% response rate for household interviews and 95% response rate for female interviews). This response rate 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 ratio of reproductive-aged women per household is applied uniformly across strata.

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

2. PMA2020 Nigeria Round 1

In Round 1, PMA2020 Nigeria was designed to give state level estimates for two states, Kaduna and Lagos. In Kaduna, urban-rural strata were considered. In Lagos, no urban-rural stratification was applied, since Lagos is predominantly urban.

PMA2020 Nigeria uses a two-stage cluster design. The master frame of EAs was based on the 2006 Nigerian population census. Census enumeration areas in Nigeria are on average 47 households in size. To obtain an enumeration area of approximately 200 households, a cluster of EAs was constructed–hereinafter referred to as EA cluster. An index enumeration area, along with a list of contiguous EAs and associated sampling probabilities, were provided by the National Population Commission (NPopC). Enumeration areas were combined into EA clusters and sampling probabilities were adjusted. Within sampling strata (i.e., Lagos and the urban/rural strata in Kaduna), EA clusters are selected using the PPS method.

Kaduna

According to the 2013 Nigeria Demographic and Health Survey (NDHS), the Kaduna mCPR for all women was 19.4%. Assuming a design effect of 3.0 and adjusted for 10% cumulative non-response rate, 2,219 women would be needed to estimate this proportion with a margin of error of 3.0% at the state level. The sample size requirements at a 5% margin of error for urban women in Kaduna with a mCPR at 30.0% are 1,073 and for rural women with a mCPR of 10.8% is 492. Thus, the state-level required sample size was larger than sum of required sample sizes by urban and rural area, and it was used as a target sample size in Kaduna.

Lagos

According to the 2013 Nigeria Demographic and Health Survey, the Lagos mCPR for all women was 24.8%. In Round 1, a design effect of 1.31 was applied and adjusted for 10% cumulative non-response rate, resulting in 1,155 women needed to estimate this proportion with a margin of error of 3.0% at the state level.

Table 1 shows the point estimate of mCPR, DEFF, the designated margin of error, and estimated sample size of women by state.

TABLE 1: MODERN CONTRACEPTIVE PREVALENCE RATES (MCPR), DESIGN EFFECT, AND ESTIMATED FEMALE SAMPLE SIZE NEEDED TO ACHIEVE SPECIFIED MARGIN OF ERROR

Strata	mCPR[a]	DEFF used	Margin of error	Number of women
Kaduna Rural	10.8	3.0[b]	.05	492
Kaduna Urban	30.0	3.0[b]	.05	1073
Kaduna State	19.4	3.0[b]	.03	2219
Lagos State	24.8	1.31	.03	1155

[a] Source: 2013 Nigeria DHS

[b] Estimated design effect in 2013 Nigeria DHS is greater than 3.

Assuming one eligible woman per household and take size of 35 households per EA cluster, 64 EA clusters in Kaduna and 34 EA clusters in Lagos were required (Table 2). Considering implementation factors, 66 and 37 EA clusters selected in Kaduna and Lagos, respectively.

TABLE 2 REQUIRED SAMPLE SIZES OF WOMEN, HOUSEHOLDS AND EA CLUSTERS: KADUNA AND LAGOS

State	Required				Implemented
	Number of women	Estimated WRA/HH	Number of HHs	Number of EA clusters	Number of EA clusters
Kaduna State	2219	1.00	2219	64	66
Lagos State	1155	1.00	1155	34	37

In sum, using the two-stage survey sampling procedure with the above sample sizes of EA clusters, households, and females, it was designed to estimate a state level mCPR for Kaduna at within 3% margin of error and the urban and rural strata with less than 5% error. Within Lagos, it was designed to estimate a state level mCPR within 3% margin of error.

3. PMA2020/Nigeria Round 2: updates

PMA2020/Nigeria continued in the two selected states: Kaduna and Lagos. No adjustments were made to the household sample size in Kaduna. After Round 1 in Lagos, however, a shortfall in the targeted sample size was observed, with only 771 completed female interviews. This was subsequently determined to be the result of a lower-than-expected response rate (75% among households and 89% among women). In addition, while one woman per household was assumed, it was much lower in Lagos. For Round 2, therefore, a few adjustments were made to calculate required sample size: a new DEFF estimates of 2.03 was applied; the average number of eligible women per household was assumed to be 0.90; and, the take size was raised to 40 households per EA cluster. The resulting required sample size was 1,791 women from 50 EA clusters. Considering low response rates during Round 1, a total of 52 EA clusters–37 from Round 1 and additionally selected 15–were included for implementation in Round 2 and onward.

4. PMA2020/Nigeria Round 3: updates

In Round 3 of PMA2020/Nigeria, the sample was expanded to generate a nationally representative estimate based on demand as well as a decision to allocate budgetary resources to cover more states. Considering budget implications and implementation capacity, a three-stage sampling approach was employed, sampling at least one state from within each of the six

geopolitical zones, inclusive of the North West and South West where Kaduna and Lagos are located, respectively. In the four zones where there were no previous PMA surveys (i.e., North East, North Central, South South, and South East zones), one state was selected from each zone independently using PPS sampling. Lagos covered one-third of the population of the South West zone, and no additional states were selected in that zone. The North West zone is the largest populated area covering a population of 36 million, which is about 25% of Nigeria's total population, according to the 2006 census. Therefore, beside Kaduna, one additional state was sampled using the PPS method. The final sample of seven states included Kaduna and Kano from North West (NW), Taraba from North East (NE), Nasarawa from North Central (NC), Lagos from South West (SW), River State from South South (SS) and Anambra from South East (SE) Zones.

Sampling within each of the selected states employed the two-stage cluster sampling procedure used for Kaduna and Lagos in prior rounds. The following sections describe the state-specific sample size estimation and construction of household and female post stratification weights to generate national-level estimates.

4.1. Sample Size Calculation

We calculated the sample size to provide the estimate of modern contraceptive prevalence rate (mCPR) with a margin-of-error (+/- 3 percentage points) in each of the five additional states. Table 3 shows the number of women necessary to estimate state level mCPR in the five states with a 3.0% margin of error. The sample sizes for Kaduna and Lagos remained as in Round 2.

TABLE 3 MODERN CONTRACEPTIVE PREVALENCE RATES (MCPR), DESIGN EFFECT, AND ESTIMATED FEMALE SAMPLE SIZE NEEDED TO ACHIEVE SPECIFIED MARGIN OF ERROR

State (Zone)	mCPR[a]	DEFF used	Margin of error	Number of women
Anambra (SE)	15.0	1.54	.03	929
Kano (NW)	4.1[b]	1.96	.03	364
Nasarawa (NC)	12.4	3.00[c]	.03	1541
Rivers (SS)	18.8	2.06	.03	1487
Taraba (NE)	7.1	1.83	.03	571

[a] Source: 2013 Nigeria DHS

[b] mCPR among all women in the North West zone, where Kano state is located. In Kano, mCPR among all women was 0.4%.

[c] Estimated design effect in 2013 Nigeria DHS is greater than 3.

The number of women per household varied by state, per the 2013 Nigeria DHS (Table 4). Given the very low mCPR in Kano and Taraba, very few clusters were needed. The EA sample, however, was further expanded in three states, Kano, Taraba and Anambra. This was in anticipation of potential increase in the mCPR and sample power constraints to monitor its level in later years. Thus, a total of 302 EA clusters were included in Round 3.

TABLE 4 REQUIRED SAMPLE SIZES OF WOMEN, HOUSEHOLDS AND EA CLUSTERS AT STATE LEVEL

State (Zone)	Required				Implemented
	Number of women	Estimated WRA/HH[a]	Number of HHs	Number of EA clusters	Number of EA clusters
Anambra (SE)	929	1.02	910	27	41
Kano (NW)	364	1.25	292	9	36
Nasarawa (NC)	1541	1.10	1401	41	40
Rivers (SS)	1487	0.85	1750	50	47
Taraba (NE)	571	1.36	420	12	20

[a] Source: 2013 Nigeria DHS

With the three-stage survey sampling procedure with the above sample sizes of EA clusters, households, and females, the resulting total sample size is adequate to estimate a national level mCPR at less than a 2.0% margin of error and state-level estimates at approximately or less than 3.0% in all seven states.

4.2. Weight Creation

The sample sizes varied substantially across the seven states, and the state sample sizes across the six zones were not proportional to the population distribution of the zones in the country. Therefore, in addition to constructing a standard sampling weight for the two-stage cluster sampling design within each state, post stratification adjustment was done to create a weight to generate national-level estimates.

Sampling weight within each state

As noted earlier, the EA clusters are selected with probability proportional to size using the master sampling frame – and when relevant stratified by urban-rural areas. Each EA cluster is then mapped and a census list of households is compiled for constructing the HH listing frame for the EA cluster.

Using Kaduna as an example, where 35 households are selected for interview within each EA cluster, the selection probability of the household is calculated as the probability of selecting the EA cluster times the probability of selecting the household in the EA cluster, and the design-survey weights are created as the inverse of the HH selection probability. The weights are further adjusted for non-response at the household level within the EA cluster.

With PPS sampling, the selection probability of a unit is:

$$\pi_i^{pps} = \frac{n * P_i}{\mathrm{i} \in \mathrm{u}^{\Sigma \mathrm{P_i}}}$$

Where *n* is the sample size, and P_i is the size of measurement (population) in the unit *i*. The denominator P_i sums over all units in the sampling frame.

The selection probability of HH in the *i-th* EA cluster is:

$$H_i = \frac{35}{P_i}$$

The design-weight is then:

$$w_i = \frac{1}{\pi_i^{pps} * H_i}$$

The design-weight for the i-th EA cluster is adjusted for (unit) non-response rate by multiplying the weight w_i by the factor R_i:

$$R_i = \frac{1}{1 - f_i}$$

where f_i is the non-response rate.

As all females age 15-49 within a household are surveyed, no further selection probability of eligible women is needed to generate the female weight (i.e., it is self-weighted). Instead the household weight is adjusted for unit non-response of the female interview at the EA level. The household and female weights are then normalized at the state level.

Post-stratification weights for national estimates[i]

Because the state-specific sample size representing each zone is not proportional to the population distribution of the zone in the country (Table 5), we applied post-stratification weights. These weights were adjusted for the over and under sampling of the seven state populations in the PMA2020 survey in relationship to the actual population distribution across the six zones in the country. We used the census data updated for the most recent period by the National Population Commission, Nigeria (http://www.population.gov.ng/index.php/state-population) to derive the population distribution in the states and zones.

TABLE 5 ZONAL CENSUS AND PMA2020 SAMPLE DISTRIBUTION

Zone	% of total population*	PMA sample (Round 3)			
		State representing the zone	Number of women	% women by zone out of total sample size (unadjusted distribution)	% of women by zone, adjusted for post-stratification weight
North West (NW)	25.6	Kaduna and Kano	10,818	42.6	25.8
North East (NE)	13.5	Taraba	2,059	8.0	13.5
North Central (NC)	14.5	Nasarawa	3,812	15.3	14.5
South West (SW)	19.7	Lagos	3,047	12.2	19.7
South South (SS)	15.0	Rivers	2,697	10.9	15.0
South East (SE)	11.7	Anambra	2,856	11.0	11.7

*Source: National Population Commission, Nigeria (http://www.population.gov.ng/index.php/state-population

Additionally, in the North West zone, we adjusted the sample distribution of Kaduna and Kano within the zone to reflect the relative population size of each state. We assumed that Kano and Kaduna were representative of the Northwest zone and the sum of their respective populations can be used to represent the zone; this made it necessary to adjust the distribution of the PMA sample to reflect the underlying state population distributions.

TABLE 6 POPULATION AND PMA SAMPLE DISTRIBUTION: KADUNA AND KANO IN NORTH WEST

State	State population*	% state population out of total population in the two states	PMA sample (Round 3)		
			Number of women	% women by state out of total sample size in the zone (unadjusted distribution)	% of women by state, adjusted for post-stratification weight
Kaduna	6,113,503	39.4	6540	60.8	39.4
Kano	9,401,288	60.6	4278	39.2	60.6

*Source: National Population Commission, Nigeria (http://www.population.gov.ng/index.php/state-population

Post-stratification weights were calculated with the following formula:

$$W_{ij} = \frac{P_j * f_{ij}}{S_i}$$

where P_j is the population fraction of the *j-th* zone in the country; S_i is the sample fraction of the *i-th* state; and f_{ij} is the fractional representation of the *i-th* state population among all states' populations sampled in the *j-th* zone. For five zones, only one state was included and thus f_{ij}=1.

The final weights for national-level estimates are then a product of state-level weights (as detailed above for both the sample design weight within the state and non-response weight) and the post-stratification weight (W_{ij}).

---

[i] As a part of a sensitivity analysis of our national level estimate, we also applied an alternative estimation procedure where the zonal stratification was ignored. In this approach, the state level population was directly used as in the weight formula:

$$W_i = \frac{\mathrm{i} \in \mathrm{u}^{\Sigma \mathrm{P_i}}}{7 * P_i}$$

where the total national-level population (sum of population from all the states) was divided by the number of states in the PMA2020 sample (n=7) multiplied by the population size of the selected state (where W_i is the weight for the *i-th* selected state in representing the fraction of the population in the country in the PMA2020 sample). The mCPR estimates were nearly identical (all women mCPR of 14.41% with this alternative method, compared to 14.43% with the method described in the memo).