**THE FEDERAL REPUBLIC
OF NIGERIA**

NATIONAL BUREAU OF STATISTICS

THE WORLD BANK
IBRD · IDA | WORLD BANK GROUP

Basic Information Document

*Nigeria Living Standards Survey
2018/19*

June 2020

ACRONYMS

CAPI	Computer Assisted Personal Interviewing
CPI	Consumer Price Index
DECDG	Development Economics Data Group
EA	Enumeration Area
FCT, Abuja	Federal Capital Territory, Abuja
GHS-Panel	General Household Survey - Panel
GPS	Global Positioning System
ICT	Information Communication Technology
HNLSS	Harmonized Nigeria Living Standards Survey
LGA	Local Government Area
LSMS	Living Standards Measurement Study
NASSCO	National Social Safety Nets Programme Office
NBS	National Bureau of Statistics
NISH	Nigeria Integrated Survey of Households
NLSS	Nigeria Living Standards Survey
NPopC	National Population Commission
SDGs	Sustainable Development Goals
TOT	Training of Trainers
TOI	Training of Interviewers
WB	World Bank

Table of Contents

Acknowledgements

1.0 Introduction

Between September of 2018 and September of 2019, the National Bureau of Statistics of Nigeria (NBS) collected the household data for the Nigeria Living Standards Survey (NLSS). The 2018/19 NLSS is the first large scale household survey in a decade, focusing on measuring living conditions of the population. The main objectives of the 2018/19 NLSS are: i) to provide critical information for production of a wide range of socio-economic and demographic indicators, including for benchmarking and monitoring of SDGs; ii) to monitor progress in population's welfare; iii) to provide statistical evidence and measure the impact on households of current and anticipated government policies. In addition, the 2018/19 NLSS could be utilized to improve other non-survey statistical information, e.g. to determine and calibrate the contribution of final consumption expenditures of households to GDP; to update the weights and determine the basket for the national Consumer Price Index (CPI); to improve the methodology and dissemination of micro-economic and welfare statistics in Nigeria.

The 2018/19 NLSS collected a comprehensive and diverse set of socio-economic and demographic data pertaining to the basic needs and conditions under which households live on a day to day basis. The 2018/19 NLSS questionnaire includes wide-ranging modules, covering demographic indicators, education, health, labour, expenditures on food and non-food goods, non-farm enterprises, household assets and durables, access to safety nets, housing conditions, economic shocks, exposure to crime and farm production indicators.

Several important methodological improvements were implemented in the 2018/19 NLSS compared to the last official survey from 2009/10. First, the 2018/19 NLSS was designed and collected using Computer Assisted Personal Interviewing (CAPI) software on tablet devices. This significantly improved the quality and speed of data collection and processing. Second, by administering a seven-day recall food consumption module instead of a month-long diary (as was done previously), the survey improved the accuracy and cost-effectiveness of collecting food expenditures data. In addition, extra effort was made to measure food quantities where households used non-standard units of measurement, resulting in better quantification of food consumption.

These improvements in questionnaire design and survey implementation techniques have produced overall high-quality data. With the support of the World Bank, the NBS adopted the best practices in the collection and processing of the 2018/19 NLSS household data. As is often the case, making such improvements and achieving high standards in survey implementation often comes at the expense of comparability with past survey rounds. Recognizing the benefits that come with a better survey infrastructure, a deliberate decision was made to use the 2018/19 NLSS as a new and improved baseline for Nigeria's living standards survey system.

This basic information document will provide all details related to the implementation and dissemination of the 2018/19 NLSS.

2.0   The Survey Instruments

Two sets of questionnaires – household and community – were used to collect information in the NLSS2018/19. The *Household Questionnaire* was administered to all households in the sample. The *Community Questionnaire* was administered to the community to collect information on the socio-economic indicators of the enumeration areas where the sample households reside.[1]

Household Questionnaire: The Household Questionnaire provides information on demographics; education; health; labour; food and non-food expenditure; household nonfarm income-generating activities; food security and shocks; safety nets; housing conditions; assets; information and communication technology; agriculture and land tenure; and other sources of household income.

Community Questionnaire: The Community Questionnaire solicits information on access to transported and infrastructure; community organizations; resource management; changes in the community; key events; community needs, actions and achievements; and local retail price information.

The contents of each questionnaire for the 2018/19 NLSS are outlined below.

Table 2-1: NLSS Household Questionnaire

Section	Topic	Respondent	Description
Cover	Cover	Field staff	Household identifiers, enumerator and supervisor identifiers, date and time of interview, questions to determine agricultural households, and observation notes by enumerator regarding the interview
1	Household Roster	All individuals	Roster of individuals living in the household, relationship to the household head, gender, year of birth, age, marital status, spouse identification, religion, parental status, date and reason of joining household if new, migration
2	Education	Individuals 3 years and above	Educational attainment, school characteristics, expenditures, repetition
3	Health	All individuals	General health status, health care utilization and cost of health services, functioning and disability status, anthropometrics and use of bed net
4	Labour	Individuals 5 years and above	Labour market participation during the last 7 days, temporary absence, job search, and wage work (includes benefits, time of work, and payment)

---

[1] The Community Questionnaire does not collect information from communities in the sociological sense. The data cannot be used to represent communities in Nigeria. The data collected at the community level represent information that is common to the households selected for inclusion in the selected sample enumeration areas (EAs).

Section	Topic	Respondent	Description
5	Remittances	Individuals 10 years and above	Remittances received (from abroad and from within Nigeria)
6A	Meals Away from Home	Most knowledgeable person	Value of food consumed outside the home during the last 7 days
6B	Food Expenditure	Person responsible for food preparations or food purchases	Quantity and value of food consumed within the household during the last 7 days, and quantity and value of food purchased during the last 30 days
6C	Aggregate Food Consumption	Person responsible for food preparations or food purchases	Number of days the household consumed each food group item and meal sharing with other household members during the last 7 days
7	Non-food Expenditure	Person responsible for household purchases	Non-food expenditure during the last week/last month/last 6 months/last 12 months
8	Food Security	Person responsible for food preparations or food purchases	Food security status of the household during the last 7 days/last 12 months
9	Non-farm Enterprises and Income Generating Activities	Owner or manager of enterprises	Roster of non-farm enterprises/activities run by members of the household, characteristics and operating status, loan, stock, costs and profit, constraints
10	Household Assets	Household head or most knowledgeable person	Ownership of assets and value
11	Credit	Household head or most knowledgeable person	Credit history including loans received, loans pending, or loan refusals
12A	Beta Don Come	Household head or most knowledgeable person	Household awareness of the Beta don come program and participation thereof
12	Safety Nets	Household head or most knowledgeable person	Household access to and utilization of safety nets besides beta don come
13	Other Household Income	Household head or most knowledgeable person	Miscellaneous income received by the household
14	Housing	Household head or most knowledgeable person	Homeownership and characteristics of home (type of roof, floor, outside wall, number of rooms, type of cookstove and cooking fuel, electricity connection, drinking water source, sanitation facility, refuse collection)
16	Economic Shocks	Household head or most knowledgeable person	Economic shocks affecting the household and coping mechanism adopted by the household
17	Crime and Security	Household head or most knowledgeable person	Key crime and security issues that household members have faced since 2016

Section	Topic	Respondent	Description
18	Agriculture	Household head or most knowledgeable person	Agriculture activities of the household, including crops, livestock and fisheries (capture and aquaculture)
19	Land Tenure	Household head or most knowledgeable person	Land tenure (agricultural and non-agricultural) status of the household, including land title registration

Table 2-2: NLSS Community Questionnaire

Section	Topic	Respondent	Description
Cover	Cover	Field staff	Community identifier variables, enumerator and supervisor identifiers, date and time of interview
C1	Respondents Characteristics	Community Focus Group	Respondent characteristics. Includes position in the community, education level and contact details
C2	Community Infrastructure and Transportation	Community Focus Group	Infrastructure availability and means of transportation to the infrastructure in the community
C3	Community Organizations	Community Focus Group	Characteristics of organizations in the community
C4	Land Prices and Credit	Community Focus Group	Land acquisition, land prices and credit access for members in the community
C5	Community Changes	Community Focus Group	Changes in the community compared to 3 years ago
C6	Conflict	Community Focus Group	Key crime and security issues that community or members of the community has faced since 2016
C8	Food Prices	Market Food Sellers	Prices of food items in the community

3.0   Sample and Weights

The 2018/19 NLSS sample is designed to provide representative estimates for the 36 states and the Federal Capital Territory (FCT), Abuja. By extension. The sample is also representative at the national and zonal levels. Although the sample is not explicitly stratified by urban and rural areas, it is possible to obtain urban and rural estimates from the NLSS data at the national level. At all stages, the relative proportion of urban and rural EAs as has been maintained

Before designing the sample for the 2018/19 NLSS, the results from the 2009/10 HNLSS were analysed to extract the sampling properties (variance, design effect, etc.) and estimate the required sample size to reach a desired precision for poverty estimates in the 2018/19 NLSS. A detailed report on this analysis is provided in Appendix A2.3. Several different sample size and allocation scenarios were considered and weighed against each other. Based on considerations of the survey budget, data quality and required precision of the estimates, it was decided to select 60 EAs per state, with 10 sampled households per EA for the 2018/19 NLSS. This results in an initial sample of 600 households per state and 22,200 households overall.

3.1 EA Selection

The sampling frame for the 2018/19 NLSS was based on the national master sample developed by the NBS, referred to as the NISH2[2] (Nigeria Integrated Survey of Households 2). This master sample was based on the enumeration areas (EAs) defined for the 2006 Nigeria Census Housing and Population conducted by National Population Commission (NPopC). The NISH2 was developed by the NBS to use as a frame for surveys with state-level domains. NISH2 EAs were drawn from another master sample that NBS developed for surveys with LGA-level domains (referred to as the "LGA master sample"). The NISH2 contains 200 EAs per state composed of 20 replicates of 10 sample EAs for each state, selected systematically from the full LGA master sample. Full details on the LGA master sample and NISH2 are provided in Appendix 2. Since the 2018/19 NLSS required domains at the state-level, the NISH2 served as the sampling frame for the survey.

Since the NISH2 is composed of state-level replicates of 10 sample EAs, a total of 6 replicates were selected from the NISH2 for each state to provide a total sample of 60 EAs per state. The 6 replicates selected for the 2018/19 NLSS in each state were selected using random systematic sampling. This sampling procedure provides a similar distribution of the sample EAs within each state as if one systematic sample of 60 EAs had been selected directly from the census frame of EAs.

Although a sample was initially drawn for Borno state, the ongoing insurgency in the state presented severe challenges in conducting the survey there. Box 1 below describes the special circumstances of the Borno sample.

---

[2] The NISH2 is distinct from the NISH which is another master sample for state domain surveys.

Box 1: Borno Sample and Weights

At the time of the 2018/19 NLSS data collection, there still was an ongoing insurgency in portions of the north east of the country, concentrated in the state of Borno. The situation in the state made it impossible for the field teams to reach large areas of the state without compromising their safety. Given this limitation it was clear that a representative sample for Borno was not possible. However, it was decided to proceed with conducting the survey in areas that the teams could access in order to collect some information on the parts of the state that were accessible.

The limited area that field staff could safely operate in in Borno necessitated an alternative sample selection process from the other states. The EA selection occurred in several stages. Initially, an attempt was made to limit the frame to selected LGAs that were considered accessible. However, after selection of the EAs from the identified LGAs, it was reported by the NBS listing teams that a large share of the selected EAs were not safe for them to visit. Therefore, an alternative approach was adopted that would better ensure the safety of the field team but compromise further the representativeness of the sample. First, the list of 788 EAs in the LGA master sample for Borno were reviewed by NBS staff in Borno and the EAs they deemed accessible were identified. The team identified 359 EAs (46%) that were accessible. These 359 EAs served as the frame for the Borno sample and 60 EAs were randomly selected from this frame. However, throughout the course of the NLSS fieldwork, additional insurgency related events occurred which resulted in 7 of the 60 EAs being inaccessible when they were to be visited. Unlike for the main sample, these EAs were not replaced. Therefore, 53 EAs were ultimately covered from the Borno sample. The listing and household selection process that followed was the same as for the rest of the states.

As a result of the different frame and selection process, the Borno household weight calculation was also distinct from the other states. The base weights were calculated by taking the inverse of the following formula:

$$p_{Li}^{Borno} = \frac{30}{N_L} \times \frac{60}{n^{acc}} \times \frac{53}{60} \times \frac{10}{M'_i} \times \frac{m'_i}{10}$$

Where:

p_{Li}^{Borno} = the overall selection probability for EA i in LGA L
N_L = the number of EAs in the 2006 census frame for LGA L
n^{acc} = number of accessible EAs in the NISH (359)
M'_i = number of households listed in the ith sample EA
m'_i = number of sample households with completed interviews in the ith sample EA

The first term represents the probability of selection into the LGA master sample, the second term is the probability of selection from among the accessible EAs in the LGA master sample, the third term is an adjustment to account for the 7 of the selected EAs that were inaccessible and not covered, the fourth term is the probability of selection of the household from the EA, and the last term is the nonresponse adjustment. Ideally, additional adjustments could have been made to more precisely reflect the under coverage of inaccessible areas but determining accessibility in the full 2006 census EA frame was not possible. Likewise, since the underlying population that the sample is representative of is not clearly defined, no calibration was performed on the Borno weights.

As a result of these marked differences in the sampling and weighting, the Borno sample is not representative of the state and must be treated separately from the data of the other states.

A fresh listing of households was conducted in the EAs selected for the 2018/19 NLSS. See Section 4.0 for details on the listing. Throughout the course of the listing, 139 of the selected EAs (or about 6%) were not able to be listed by the field teams. The primary reason the teams were not able to conduct the listing in these EAs was due to security issues in the country. The fieldwork period of the 2018/19 NLSS saw events related to the insurgency in the north east of the country, clashes between farmers and herdsman, and roving groups of bandits. These events made it impossible for the interviewers to visit the EAs in the villages and areas affected by these conflict events. In addition to security issues, some EAs had been demolished or abandoned since the 2006 census was conducted. In order to not compromise the sample size and thus the statistical power of the estimates, it was decided to replace these 139 EAs. Additional EAs from the same state and sector were randomly selected from the remaining NISH2 EAs to replace each EA that could not be listed by the field teams. This necessary exclusion of conflict affected areas implies that the sample is representative of areas of Nigeria that were accessible during the 2018/19 NLSS fieldwork period. The sample will not reflect conditions in areas that were undergoing conflict at that time. This compromise was necessary to ensure the safety of interviewers.

3.2 Household Selection

Following the listing, the 10 households to be interviewed were selected from the listed households. These households were selected systemically after sorting by the order in which the households were listed. This systematic sampling helped to ensure that the selected households were well dispersed across the EA and thereby limit the potential for clustering of the selected households within an EA.

Occasionally, interviewers would encounter selected households that were not able to be interviewed (e.g. due to migration, refusal, etc.). In order to preserve the sample size and statistical power, households that could not be interviewed were replaced with an additional randomly selected household from the EA. Replacement households had to be requested by the field teams on a case-by-case basis and the replacement household was sent by the CAPI managers from NBS headquarters. Interviewers were required to submit a record for each household that was replaced, and justification given for their replacement. These replaced households are included in the disseminated data. However, replacements were relatively rare with only 2% of sampled households not able to be interviewed and replaced. Table 3-1 presents the final distribution of successfully and unsuccessfully interviewed households. The unsuccessful interviews also included partially completed interviews that were not replaced.

Table 3-2 present the reasons why households could not be interviewed. Refusals were very few with only 48 households refusing to participate in the interview. The primary reasons for nonresponse was that the household was unavailable for an extended period or had moved away from the EA. Partial interviews were also very rare.

Table 3-1: Final Sample Composition		
State	**Fully Interviewed households**	**Households not able to be fully interviewed**
Abia	600	6
Adamawa	600	1
Akwa Ibom	600	3
Anambra	600	5
Bauchi	600	9
Bayelsa	600	2
Benue	600	1
Borno	530	1
Cross River	600	2
Delta	600	11
Ebonyi	600	15
Edo	600	27
Ekiti	599	8
Enugu	600	36
Gombe	597	7
Imo	600	24
Jigawa	599	9
Kaduna	600	12
Kano	600	1
Katsina	600	0
Kebbi	600	6
Kogi	599	24
Kwara	599	7
Lagos	600	41
Nasarawa	600	5
Niger	600	10
Ogun	600	27
Ondo	600	11
Osun	598	35
Oyo	599	27
Plateau	600	5
Rivers	600	56
Sokoto	600	6
Taraba	599	3
Yobe	600	0
Zamfara	600	0
FCT	599	26
Total	**22,118**	**469**

Table 3-2: Reasons for Nonresponse	
Reason for nonresponse	# of households
Partially Complete (Refused)	6
Partially Complete (Unavailable)	2
Unable To Identify Household	91
Household Has Moved Away	116
Long Term Unavailable	206
Refused	48

3.3 Survey Weights

In order for the sample estimates from the 2018/19 NLSS data to be representative of the population, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample household is equal to the inverse of its probability of selection (calculated by multiplying the probabilities at each sampling stage). The probability of selection for the 2018/19 NLSS sample households can be defined as follows:

From

$$p_{SLi} = \frac{60}{N_s} \times \frac{10}{M'_{SLi}}$$

Where:

p_{SLi} = overall sampling probability for households selected based on the NISH2 master sample in the ith sample EA in LGA L within state S

N_S = total number of EAs in the 2006 Census frame for state S

M'_{SLi} = number of households listed in the ith sample EA in LGA L of state S

The first component is the probability of selection of the EA while the second is the probability of selection of the household within the EA. The EA selection probability is a combination of (1) the selection probability into the LGA master sample from the 2006 census frame, (2) the selection probability into the NISH2 from the LGA master sample, and (3) selection probability from the NISH2 to the 2018/19 NLSS sample EAs. The simple form of this equation is a result of the selection process for the NISH2 which results in this equal EA selection probability within each state. The full derivation of the above equation is provided in Appendix A2.5.

The basic weight for the 2018/19 NLSS sample households is the inverse of this probability of selection, calculated as follows:

$$W_{SLi} = \frac{1}{p_{SLi}}$$

where:

W_{SLi} =   basic weight for the 2018/19 NLSS sample households in the ith sample EA in LGA L of state S

It is also important to adjust the weights to take into account the non-interview households in each sample EA. The adjustment for nonresponse is conducted at the EA level. The adjusted base weight (W'_{SLi}) for the sample households in the ith sample EA in LGA L of state S can be expressed as follows:

$$W'_{SLi} = W_{SLi} \times \frac{10}{m'_{SLi}} \, ,$$

where:

m'_{SLi} =   number of sample households with completed interviews in the ith sample EA in state S of LGA L

Note that only households that were *fully* interviewed were included in m'_{SLi}. Households with partial interviews were excluded when calculating the weights. Though the information collected from these households remains in the public data sets, no weight was calculated for these households.

3.3.1   Calibration

The base weight obtained above reflects the sample selection process in its entirety. However, in order to better ensure the weighted estimates obtained from the survey truly reflect the underlying population distribution across the strata, the weights can be calibrated to known population totals. Following the base weight calculation above, the 2018/19 NLSS weights were calibrated to projected total state population in 2019[3] using a generalized regression approach. The calibration adjustment was performed at the EA level such that all households within the EA maintain the same weight.

The household weights can be found in the cover page data file (***secta_cover.dta***)*.* The variable name in the data file is ***wt_final***.

---

[3] The National Population Commission is currently revising their state population projections. Once the revised projections are officially released, the NLSS weights will be adjusted and calibrated to the new projections.

4.0   Listing

Prior to selection of the households to be interviewed, a fresh listing of all households contained within each EA was conducted. The household listing was performed quarterly (every 3 months) across the 36 states of the country plus the FCT. The listing was done quarterly to limit the amount of time between the listing and the main fieldwork activities. The 60 EAs in each state were systematically sorted into 4 separate groups of 15 EAs to be listed in each quarter. The selection was done systematically in order to ensure that the EAs included in each quarter covered both rural and urban areas as well as covered as many LGAs as possible.

A separate team of interviewers conducted the listing. Each state had one listing team consisting of a "lister" who enumerated the households and a "mapper" who drew a sketch map of each EA demarcating the boundaries and all households listed within. These sketch maps were provided to the main NLSS fieldwork teams to facilitate locating and identification of the selected households. Two trainings were conducted for the listing team. The main training was given prior to the first quarter listing and a refresher training was also given prior to the third quarter listing.

A listing monitoring team consisting of personnel from NBS headquarters was established and trained along with the listing team. In addition to the main and refresher trainings with the listing teams, the listing monitors were also given a short refresher training prior to the second and fourth quarter listing. The monitors then facilitated a similar refresher training to the listing teams in the states in the second and fourth quarter listing. The monitoring team observed listing fieldwork in all four quarters, identified any issues and guided the teams on how to improve.

The household listing was conducted using Survey Solutions CAPI, with inbuilt GPS location collection. Using this feature, interviewers collected the GPS location of all buildings enumerated within the EA. Although digitized and georeferenced EA maps were not available, real-time access to the listing data and geolocations of the EAs and buildings allowed for monitoring of the incoming data so issues of undercoverage of an EA could be identified and rectified. The GPS information from the listing exercise also helped to build quality control checks for the main survey team during the data collection exercise.

The first quarter household listing exercise was undertaken in June/July 2018. Due to delays in the commencing the main survey data collection, the second quarter listing was undertaken in December 2018, while the third and fourth quarter listing exercise occurred in March and June 2019 respectively. Following the completion of each quarter's listing, the households to be interviewed by the main survey team were selected. The small gap in time between the listing and fieldwork helped to limit the possibility that a listed household had moved or disbanded between the listing and the main field interview.

5.0 Pre-test and Pilot

5.1 Pre-test

A pre-test was conducted in December 2016 in one urban and one rural EA within the FCT. The purpose of the pre-test was to test the questionnaire and CAPI application to identify any major issues and areas to improve before conducting a full-scale pilot. A total of 26 households were interviewed by a small team of interviewers.

Due to logistical challenges, the main survey activities (including pilot) did not follow immediately after the pre-test. The duration between the pre-test and the pilot was one and half years.

5.2 Pilot

As part of the questionnaire design and quality control process, a pilot exercise was conducted for the 2018/19 NLSS. The pilot fieldwork covered the period from June 19 – 24, 2018.

Prior to the pilot exercise, a six-day training was organized to train interviewers, supervisors and other personnel. The training had both in-class and field practice component. Participants were taken through the various contents of the questionnaires, with detailed explanations of concepts and mock interviews. After the in-class training, a field practice was conducted in 4 nearby urban EAs, followed by a debriefing. The field practice gave the participants first-hand field experience that was vital for the main pilot fieldwork.

The pilot fieldwork occurred in 4 states – Delta, Kogi, Bauchi and Lagos – over a period of 6 days. In each of these states, 5 EAs were visited and 5 households sampled from each EA. The households interviewed came from a 2014 household listing exercise conducted by the bureau. The EAs included in the pilot exercise in these states were not part of the main sample EAs. Given the security situation in some of the states, not all the EAs selected for the pilot were visited by respective teams. Specifically, during the field operations in Kogi state, the field team's vehicle was shot at in a suspected kidnapping attempt, so the pilot exercise in that state had to be suspended. In the end, 74 households were interviewed during the pilot.

As a result of the lessons learned from the pilot, the questionnaire was reviewed, with questions and sections removed in order to shorten the length of interview. Main fieldwork plans and other logistics were revisited as a result of the feedback lessons from the pilot exercise.

6.0   Training of Field Staff and Data Quality Monitoring Team

6.1 Main Survey Training Design

Two levels of training were implemented. The first level training (training of trainers (ToT)) was organized at the NBS headquarters in Abuja. The participants of the first level training included 24 senior technical staff from various departments within the bureau, 24 data editors and 7 data managers (from the ICT department), as well as 6 coordinators, all based at the NBS headquarters. The ToT focused on equipping participants with detailed technical understanding of the survey instruments, including CAPI and field practice, so that they can train the field personnel in the second level training. At the end of the ToT, the technical staff and the data editors were examined on different aspects of the questionnaire and the survey, and training and presentation skills. Selection of the senior technical staff to be trainers in the second level training was based on performance in the exam, in-class performance, training and presentation skills, and the individual's ability to explain difficult concepts and manage a class. Overall, 18 senior technical staff were selected as trainers. The selection of the 18 data editors, however, was based only on performance in the exam. The ToT lasted for 10 days.

The second level training (training of interviewers (ToI)) was organized in Keffi, Nassarawa state. The participants for the ToI comprised of NBS personnel from the states, including 6 zonal controllers, 37 state officers, 37 supervisors, and 148 interviewers. The trainers, data editors, coordinators and other stakeholders also participated in this training. The ToI lasted for 15 days, focusing on providing participants with detailed understanding of the survey and CAPI, elucidation of important concepts and questions in the questionnaire, in-class exercises, mock interviews, field practice and evaluation. At the end of the training, participants were examined, and the best candidates (interviewers and supervisors), selected to participate in the main fieldwork that lasted 12 months.

The training was given to the field staff by the trainers and data editors/managers who were selected from the first level training. Due to the large number of participants, three training rooms were used for the training, where about five or six resource persons conducted the training in each room.

The World Bank technical missions also provided support during the first and second level trainings. As an additional means of quality control, NBS hired 3 consultants with different expertise to participate in the second level training. The National Social Safety Nets Programme Office (NASSCO) also sent some of their staff to participate in the second level training.

7.0   Field Work

7.1 Organization of Fieldwork

Each state had one field team comprising one supervisor and three interviewers, who worked in a roving manner. The team traveled to an EA, interviewed all selected households and conducted the community interview, including collecting market prices, and then moved to the next EA. Teams spent on average 3 days in an EA, with each interviewer interviewing one household per day. The supervisor administered the community questionnaire, collected the market prices, and then interviewed one remaining household after the 3 interviewers had each interviewed 3 households over the 3-day period. Thus, in addition to the community questionnaire, the supervisor conducted one household interview per EA. Besides interviewing all selected households fully and accurately, spending 3 days per EA allowed the team to also address all error checks, comments, and feedback that were sent to them by the data editor, NBS headquarters team, and the World Bank technical team.

Given that the survey was conducted over a 12-month period, the teams were in the field throughout the duration of the fieldwork except for scheduled breaks. Longer breaks in the fieldwork occurred in the holiday period of late December/early January as well as during the federal elections held in February 2019. Shorter breaks occurred for other holidays such as Easter and Eid. Figure 7-1 shows the number of interviews conducted by month throughout the survey period. The largest deviation from the average occurred in February when the fieldwork was suspended around the time of the federal elections. Table 7-1 also presents the number of households interviewed in each state across the fieldwork months. Note that the start date of fieldwork was September 28, 2018 though around half of the states started their fieldwork on October 2, 2018. These staggered start dates were implemented so that the monitors (who were assigned to 2 states) could be present for the start of fieldwork in every state.
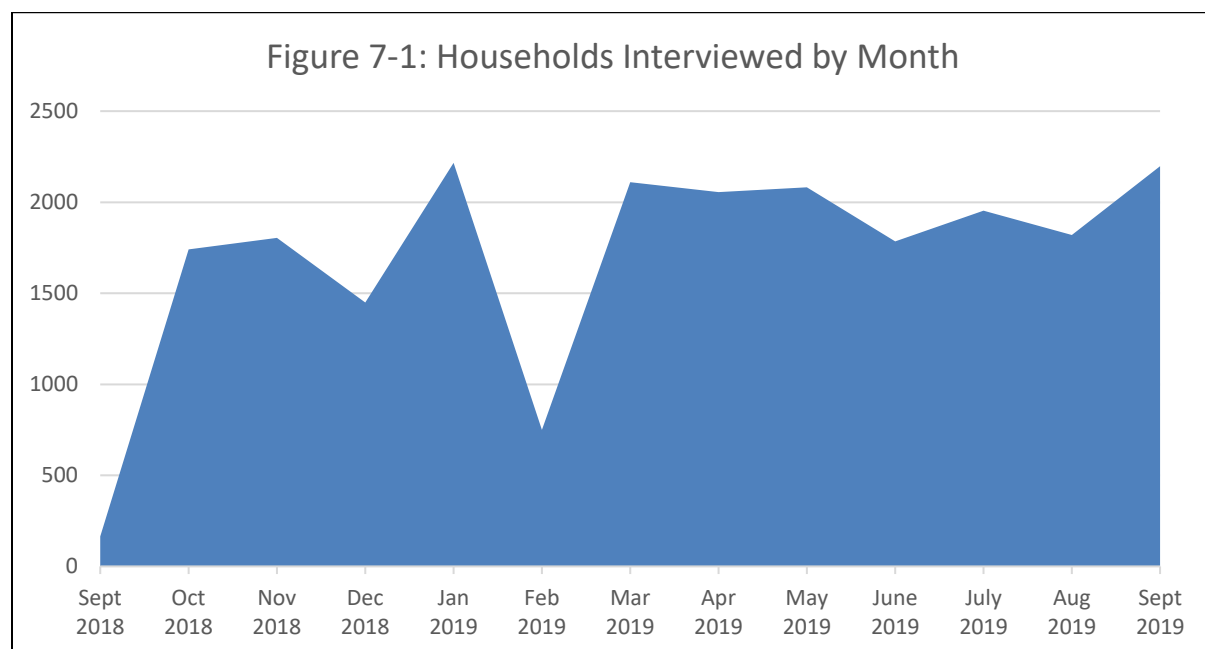

Figure 7-1: Households Interviewed by Month

Table 7-1: Households Interviewed by Month

State	Sept 2018	Oct 2018	Nov 2018	Dec 2018	Jan 2019	Feb 2019	Mar 2019	Apr 2019	May 2019	June 2019	July 2019	Aug 2019	Sept 2019
Abia		50	50	40	60	20	59	55	53	50	57	46	60
Adamawa	10	30	50	30	60	30	57	56	51	50	66	50	60
Akwa Ibom		50	50	40	60	20	57	57	56	48	56	46	60
Anambra		50	50	40	60	20	57	57	53	50	57	46	60
Bauchi		50	50	40	60	20	50	60	59	45	56	46	64
Bayelsa		50	50	40	60	20	56	57	55	49	50	53	60
Benue	10	48	50	32	60	20	54	60	56	50	50	50	60
Borno		50	43	46	61	20	58	52	60	20	39	51	30
Cross River	10	47	43	40	60	20	60	54	56	49	51	50	60
Delta		50	50	40	60	20	55	59	53	50	57	46	60
Ebonyi	10	48	42	40	60	20	60	54	56	47	53	50	60
Edo	9	49	46	36	60	20	60	52	56	48	51	52	61
Ekiti	7	52	40	40	60	20	60	55	53	52	55	45	60
Enugu		50	49	41	60	20	57	57	56	47	53	50	60
Gombe	10	40	50	40	57	20	57	57	56	50	50	50	60
Imo	9	48	47	36	60	20	54	59	57	47	53	50	60
Jigawa		50	49	41	60	20	60	50	60	46	54	50	60
Kaduna	10	37	57	36	60	20	56	54	57	52	48	53	60
Kano	10	39	54	37	60	20	60	50	57	46	57	50	60
Katsina		50	47	43	60	20	50	60	60	48	56	46	60
Kebbi		50	47	43	60	20	56	57	57	47	53	50	60
Kogi	10	40	55	35	60	20	57	56	57	48	52	50	60
Kwara	10	39	54	36	60	20	57	57	56	47	53	50	60
Lagos	7	44	53	37	58	21	57	56	56	48	53	50	60
Nasarawa		50	50	40	60	20	57	53	59	50	55	44	62
Niger		50	47	43	60	20	60	54	53	53	47	53	60
Ogun		50	50	40	60	20	60	54	56	50	50	50	60
Ondo		50	47	43	60	20	60	54	56	50	50	50	60
Osun	7	42	53	37	60	20	57	53	59	48	53	49	61
Oyo		50	50	39	60	20	58	57	56	49	52	50	60
Plateau		50	48	42	60	20	58	55	55	52	54	46	60
Rivers	7	43	53	37	60	20	57	57	56	47	53	50	60
Sokoto	4	50	44	42	60	20	58	52	56	54	50	50	60
Taraba		51	47	43	60	19	49	61	57	46	57	50	60
Yobe	10	47	43	40	60	20	54	56	58	51	50	51	60
Zamfara	10	50	40	40	60	20	57	53	60	50	53	47	60
FCT	4	46	56	34	60	20	60	56	54	50	50	50	60
Total	**164**	**1740**	**1804**	**1449**	**2216**	**750**	**2109**	**2056**	**2081**	**1784**	**1954**	**1820**	**2198**

Teams were assigned to either of 3 fieldwork schedules which specified which EA should be covered and included staggered breaks across the survey teams. This schedule was duly followed across the different management and technical structures of the project. Table 7-2 below shows the number of EAs that an assigned team was expected to complete before taking first long break, assuming a 31-day calendar month. For instance, teams operating with schedule 1 took their first long break after they had completed three EAs, while those using schedule 2 and 3 did so after completing 2 and 4 EAs respectively. After the first long break, all subsequent long breaks occurred after each 3 EAs. Furthermore, the order in which the EAs were covered was predetermined and randomized in order to provide more uniform coverage of the state throughout the fieldwork period.

7.2 Gift to Households

As a show of appreciation for the households' participation, all households that were interviewed, were given a gift. The gifts came in the form of food flask (container for storing food). Households were very appreciative of the gifts.

7.3 Pre-loaded Information

Basic identification information (location, household head name, phone number, etc.) on every household was pre-loaded in the CAPI assignments for each interviewer. The information was pre-loaded to assist interviewers in locating and identifying the household. The pre-loaded basic household information was derived from the household listing exercise.

7.4 In-Person Monitoring of Field work

To ensure that good quality data is collected, an extensive field monitoring exercise was mounted during the 12-months of data collection. The first monitoring was implemented immediately after the second level training and at the start of the fieldwork. In the first round of monitoring, one senior technical staff (trainer) that was part of the first and second training was assigned to monitor field teams in 2-3 states. These monitors visited their respective teams 4 times during the 12-month period, with the visits scheduled such that the monitors visit the field teams every quarter. On each visit to the team, the monitors spent up to 6 days with the team, including field visits to oversee interviews and conduct spot-checks.

The monitors were tasked with ensuring the interviewers were fully conforming with the procedures laid out in the manual and explained during training as well as effecting necessary corrections and tackling any problems that the field teams might face. During each visit to the teams, monitors were given a monitoring questionnaire to complete and upload to the project server. This was to ensure that the monitors visited the teams, spent the required number of days with the teams, and reported vividly and accurately, their observations from their monitoring visit. The monitors were also charged to continue to follow-up with the teams in their respective states throughout the course of the fieldwork, remotely address any issues or challenges that they might have, and filter unsurmountable challenges to the management and senior technical persons to address.

Table 7-2: Example Fieldwork Schedules

Day	Sched1	Sched2	Sched3
1	EA 1	EA 1	EA 1
2	EA 1	EA 1	EA 1
3	EA 1	EA 1	EA 1
4	EA 1	EA 1	EA 1
5	EA 2	EA 2	EA 2
6	EA 2	EA 2	EA 2
7	Sunday		
8	EA 2	EA 2	EA 2
9	EA 2	EA 2	EA 2
10	Break	Break	Break
11	EA 3	Break	EA 3
12	EA 3	Break	EA 3
13	EA 3	Break	EA 3
14	Sunday		
15	EA 3	Break	EA 3
16	Break	EA 3	EA 4
17	Break	EA 3	EA 4
18	Break	EA 3	EA 4
19	Break	EA 3	EA 4
20	Break	EA 4	Break
21	Sunday		
22	EA 4	EA 4	Break
23	EA 4	EA 4	Break
24	EA 4	EA 4	Break
25	EA 4	Break	Break
26	EA 5	EA 5	EA 5
27	EA 5	EA 5	EA 5
28	Sunday		
29	EA 5	EA 5	EA 5
30	EA 5	EA 5	EA 5
31	Break	Break	Break
1	EA 6	Break	EA 6
2	EA 6	Break	EA 6
3	EA 6	Break	EA 6
4	EA 6	Break	EA 6

During the periods when the monitors are not with the teams, the state officers and zonal controllers took in-person field monitoring responsibilities, reporting directly to headquarters any issues/challenges that could mar the quality of the data collected. While the state officers monitored in their own state, the zonal controllers conducted monitoring in at least 2 states (the zonal headquarters state and one other state of the same zone).

7.5 Remote Monitoring of fieldwork

In addition to the in-person monitoring of quality of the data collection by the monitors, there was also an extensive remote monitoring effort conducted by NBS ICT team and the World Bank technical team. The first level of remote monitoring was performed by the NBS ICT (data editors and data managers). The data editors would review every incoming interview from the field for any potential errors or omissions. Their review was also complimented by the second level of monitoring performed by the data managers and the World Bank technical team. Each day, the live data was downloaded from the server and a comprehensive set of error, outlier, and consistency checks were performed on newly submitted interviews. A report was generated for each case and provided to the data editors. The data editors reviewed the issues identified and included them in their own review of the interviews. If any issues were identified by the data editor's review or from the global data checks, then the data editor would make comments in the interview and reject it back to the interviewer for them to address the identified issues. Data editors would also contact interviewers or supervisors for persistent or complicated problems that need to be more thoroughly addressed. After the issues were addressed either through a re-interview of the household or explanation by the interviewer, the interviewer would send the interview back to the data editor who would either approve the case or reject back for further clarification.

In addition to the daily interview review and global data checks, a dashboard was developed which tracked fieldwork progress and interviewer performance. This dashboard allowed NBS coordinators and the World Bank technical team to more broadly monitor data quality and spot consistent issues or particular teams who needed additional attention.

8.0 Data Management and Description of Datasets

8.1 Data Management

8.1.1 CAPI

The 2018/19 NLSS was conducted using the Survey Solutions Computer Assisted Person Interview (CAPI) platform. The Survey Solutions software was developed and maintained by the Development Economics Data Group (DECDG) at the World Bank. Each interviewer and supervisor was given a tablet which they used to conduct the interviews. Overall, implementation of the survey using Survey Solutions CAPI was highly successful, as it allowed for timely availability of the data from completed interviews and real-time quality checks.

8.1.2 Data Communication System

The data communication system used in 2018/19 NLSS was highly automated. Each field team was given a mobile modem to allow for internet connectivity and daily synchronization of their assignments and completed interviews. This ensured that headquarters in Abuja had access to the data in real-time. Once each interview was completed and uploaded to the server, the data was first reviewed by the data editors. The data was also downloaded from the server, and Stata dofiles run on the downloaded data to check for additional errors that were not captured by the Survey Solutions application during data collection and entry. An excel error file is generated following the running of the Stata dofiles on the raw dataset. Information contained in the excel error files are communicated back to respective field interviewers for action by the interviewers. This action was done on a daily basis for the duration of the survey.

8.1.3 Data Cleaning

The data cleaning process was done in three main stages. The first stage was to ensure proper quality control during the fieldwork. This was achieved in part by incorporating validation and consistency checks into the Survey Solutions application used for the data collection and designed to highlight many of the errors that occurred during the fieldwork.

The second stage cleaning involved the use of data editors and data assistants. As indicated above, once the interview is completed and uploaded to the server, the data editors review completed interview for inconsistencies and extreme values. Depending on the outcome, they can either approve or reject the case. If rejected, the case goes back to the respective interviewer's tablet upon synchronization. Special care was taken to see that the households included in the data matched with the selected sample and where there were differences, these were properly assessed and documented. Additional errors observed were compiled into error reports that were regularly sent to the teams. These errors were then corrected based on re-visits to the household on the instruction of the supervisor. The data that had gone through this first stage of cleaning was then approved by the Data Editor. After the data editor's approval of the interview on Survey Solutions server, the Headquarters also reviews and depending on the outcome, can either reject or approve.

The third stage of cleaning involved a comprehensive review of the final raw data following the first and second stage cleaning. Every variable was examined individually for (1) consistency with other sections and variables, (2) out of range responses, and (3) outliers. However, special care was taken to avoid making strong assumptions when resolving potential errors. Some minor errors remain in the data where the diagnosis and/or solution were unclear to the data cleaning team.

8.2 Description of Datasets

As indicated above, the 2018/19 NLSS was implemented over a 12-month period, with different households interviewed each day in each state, except for scheduled fieldwork breaks. The household questionnaire was administered to sampled households in each EA, while the community questionnaire was administered at the level of the EA or community in which the EA is domiciled.

Table 8-1 shows the sections of the Household Questionnaire and their corresponding datasets.

Table 8-1: Household datasets

Section	Section Name	Dataset Filename
A	Cover	secta_cover
1	Household Roster	sect1_roster
2	Education	sect2_education
3	Health	sect3_health
4	Labour	sect4a1_labour
		sect4a2_labour
5	Remittances	sect5_remittances
6A	Meals Away from Home	sect6a_meal_outside
6B	Food Expenditure	sect6b_food_cons
6C	Aggregate Food Consumption	sect6c_aggregate_food_1
		sect6c_aggregate_food_2
		sect6c_aggregate_food_3
7	Non-Food Expenditures	sect07_7day
		sect07_30day
		sect07_12month
8	Food Security	sect8_food_security
9	Non-Farm Enterprise	sect9a_enterprise
		sect9b_enterprise
		sect9c_enterprise
		sect9d_enterprise
10	Household Assets	sect10_assets
11	Credit	sect11a_credit
		sect11b_credit
		sect11c_credit

12	Beta Don Come	sect12a_safety
	Social Safety Nets	sect12b_safety
13	Other Household Income	sect13_income
14	Housing	sect14_housing
16	Economic Shocks	sect16_shocks
17	Crime and Security	sect17_crime_security
18	Agriculture	sect18_agriculture
19	Land Tenure	sect19a_land
		sect19b_land

8.2.1   Community Data

Table 8-28-2 shows the sections of the Community Questionnaire and their corresponding data sets.

Table 8-2: Community datasets

Section	Section Name	Dataset Filename
C	Cover	sectc_cover
C1	Respondents Characteristics	sectc1_resp
C2	Community Infrastructure and Transportation	sectc2_infra
C3	Community Organizations	sectc3_org
C4	Land Access	sectc4a_land
	Land Prices	sectc4a_prices
	Credit	sectc4b_credit
C5	Community Changes	sectc5_changes
C6	Conflict	sectc6_conflict
C8	Food Prices	sectc8_foodprices

8.2.2   Auxiliary information

Two additional data sets are released with 2018/19 NLSS which includes auxiliary information of the interviews – *sect_aux.dta* and *sect_result.dta*. The first, *sect_aux* include the starting and the ending time of the interviews, respondents of the individual section of the questionnaire. The second file contains information on the interview completion rate, language of the interview, presence of a monitoring officer during the interview and some dwelling characteristics for households that refused to be interviewed. These two datasets also provide additional quality control information about the survey and the data collected.

8.2.3   Confidential information

Note that, for purposes of maintaining the confidentiality of the data, all names, addresses, and phone numbers have been removed from the datasets. Additionally, the GPS coordinates have also been removed as these could be used to locate households with accuracy.

8.2.4   Non-Standard Units Conversion Factors

Food quantities are often reported in non-standard units in the data. In order to convert from non-standard units to the more widely understood standard units (kilograms and litres), conversion factors are included in the food consumption datasets.

8.2.5   Consumption Aggregate

In addition to the unprocessed household and community files, a processed data file that contains the calculated consumption aggregates – *totcons.dta* – is also provided. The consumption aggregate combines different sources of household consumption and expenditure and servers as the primary measure of household material wellbeing as well as determine their poverty status. Full details of the methodology used to calculate the consumption aggregates as well as the variables contained in the *totcons.dta* file can be found in the methodological note released alongside this document.

9.0   Using the Data

9.1 File Structure

The data should always be used in conjunction with the questionnaire and the interviewer's instruction manual. Where there are no issues of confidentiality, all the variables from the questionnaire have been included in the data sets. In some cases, there is an additional variable which contains the "other specify" information that was written in the questionnaire. So, for example, if there is a variable with two parts question 10a and question 10b, a third variable, question 10c, might be added which would contain the other "specify information". In some cases, the other specify variable will be indicated with an "_os" attached to the variable name.

9.2 Merging Datasets

9.2.1   Household

All household datasets contain a variable (***hhid***), which is a unique identifier for the household. This variable is used as the unique key variable in the merging of all household type datasets. In some of the other types of datasets, additional key variables may be required in the merging process. In the case of individual type files, the variable that uniquely identifies the individual in the household is ***indiv***. In order to merge any two individual level type files, both the variables ***hhid*** and ***indiv*** should be used.

9.2.2   Community Datasets

The community questionnaire is administered at the EA level so the location variables ***lga*** for local government area (LGA) and ***ea*** are unique for each community questionnaire. An additional variable ***cluster_id*** is included in each file which is a concatenation of ***lga*** and ***ea*** and therefore uniquely identifies each EA.

9.3 Food Unit Measures

When collecting information on food quantities (e.g. amount of food consumed, etc.), respondents were allowed to report in any unit that they were most familiar with. Quite often, respondents provided quantities in non-standard units like "milk cup", "mudu", or "sack" (as opposed to standard units like kilograms, litres, etc.). In the 2018/19 NLSS, an expanded unit list was used to account for a wider range of possible units that are common in Nigeria. In addition, for some units, respondents were required to provide a size (small, medium, or large) for the unit. This element was added to better account for variations in the size of some units. In order to standardize the relative sizes of units, interviewers would show the respondent a photo of the unit including the difference sizes as applicable. The respondent would then indicate the appropriate size for the unit they are reporting in. This was particularly

important for vaguely defined units such as "piece" or "heap" which are relatively common. For these units, item-specific photos were shown to the respondent.

9.3.1   Unit Conversion Factors

In order to utilise and compare quantities in different units they must be converted into a common unit using conversion factors. As part of the General Household Survey Panel (GHS-Panel), NBS has conducted a market survey to collect conversion factors and reference photos for a wide array of food item/crop-units. This market survey was conducted in 12 states across the 6 zones in order to capture differences in units and conversion across the country. Where there were sufficient observations, zone-level conversions were calculated.

Additional items and units were identified throughout the course of the fieldwork for which no conversion factors were available. An additional effort was made by NBS to collected additional conversion factors and reference photographs for these items to ensure comprehensive coverage of reported items and units.

These conversion factors are included as variables in respective datasets to make it easier for the user. For example, the dataset **sect6b_food_cons.dta** contains question 2, which asks how much the household consumed of each food item in the last 7 days. Question 2 has four main components: the quantity (2a), the unit (2b), the size (2ac) (where applicable), and the conversion. The conversion variables all take the form of "***_cvn**". In order to convert the quantity to kilograms or litres, all the data user needs to do is multiply the quantity by the conversion. For question 2 in the example above, that would be **s06bq02a*s06bq02_cvn**.

The conversion factors included in the data files cover a majority of item/crop-unit combinations observed in the data set, there are still some gaps where conversion factors are not available. There is an ongoing effort to fill these gaps and updated conversion factors will be released as they become available.

9.3.2   Reference Photo Album

The 2018/19 NLSS relied heavily on the reference photo albums created as a result of the GHS-Panel project implemented by the bureau. In an effort to obtain conversion factors for the numerous non-standard units for food and agricultural production items, a market survey was implemented for the GHS-Panel project. During the market survey, item-unit weights were collected in a systematic manner, where interviewers were instructed to follow strict protocols when taking photographs of items, such as including a reference object (typically a standard sized bottle of water) to provide the respondent with a frame of reference for the size of the unit. For units with multiple sizes, all of the relevant sizes were taken in the same photo for easier comparison by the respondent. The reference photos taken during the market survey were compiled into an album that was printed and provided to all interviewers. Item-specific photos were included for non-container units (piece, heap, bunch, stalk) while only one photo

of containers (e.g. milk cup, tiya, mudu) were included. The reference photo album that was used by interviewers is included with the additional documentation on the website (see "Photo Aids") The procedures used for collection of the reference photos as well as the conversion factors followed the guidelines laid out in a guidebook produced by the Living Standards Measurement Study (LSMS) team, *The Use of Non-Standard Units for the Collection of Food Quantity: A Guidebook for Improving the Measurement of Food Consumption and Agricultural Production in Living Standards Surveys*.

10.0 Overall Problems and Challenges Faced

Designing and implementing a complex survey such as the NLSS presents various challenges. In this section we outline some key issues that arose, lessons learned, and make recommendations for future NLSS.

10.1   Security Problems

The most significant challenge faced during the 2018/19 NLSS was the security situation in the North East and North Central Zones of the country, particularly Borno state. In these areas, there were several areas that were impassable due to roadblocks by security forces or were deemed too hazardous for field staff to visit. Security concerns were present in other localities within and outside of the North East. For instance, during the pilot exercise, a World Bank and NBS headquarters monitoring team's vehicle was shot at while departing an EA, a suspected kidnapping attempt. Thus, sometimes the teams had to adjust their plans based on the situation on the ground or obtain security clearance from appropriate state or LGA.

Further, the general security situation in the country prevented World Bank staff and consultants from observing and monitoring fieldwork directly. As described above, a further emphasis was placed on remote monitoring of the incoming data.

10.2   Federal Elections

Due to substantial delays in the commencement of the fieldwork activities, the fieldwork period overlapped with federal elections in February 2019. Out of an abundance of caution, the fieldwork was suspended for 2 weeks before and after the day of the election. This suspension was due to serious security concerns around the election period. Further, the election was abruptly postponed by one week which resulted in an extension of the break in fieldwork. Although the election did not otherwise have a noticeable impact on the fieldwork and the respondents' willingness to participate in the survey during this period, it was nevertheless a disruption in the fieldwork schedule.

Appendix 1: How to Obtain Copies of the Data

The data are available through the NBS web site:

http://www.nigerianstat.gov.ng/nada

or through the World Bank's Microdata Library:

https://microdata.worldbank.org/

Users do not need to obtain the permission of the NBS to receive a copy of the data but will be asked to fill in a data access agreement. In this agreement, users agree to: (a) cite the National Bureau of Statistics as the collector of the data in all reports, publications and presentations; (b) provide copies of all reports publications and presentations to the National Bureau of Statistics (see address below); and (c) not pass the data to any third parties for any reasons.

Biyi Fafunmi
Head of Department - ICT
Plot 762, Independence Avenue,
Central Business District,
FCT, Abuja
Nigeria
www.nigerianstat.gov.ng/nada
Email: biyifafunmi@nigerianstat.gov.ng

Appendix 2: Sampling Details

A2.1. LGA Master Sample

The NBS developed a master sample of EAs in each Local Government Area (LGA) that could be used for any LGA-level survey. For this LGA master sample 30 EAs were selected with equal probability within each LGA for the 36 states, and 40 EAs were selected in each LGA for Abuja FCT. There are 768 LGAs in the 36 states of Nigeria and 6 LGAs in Abuja, so a total of 23,280 EAs were selected for the LGA master sample. The NBS obtained copies of the maps from the 2006 Census for these sample EAs from the National Population Commission (NPopC). The EAs selected for the master sample in each LGA were based on 3 independent replicates of 10 households each (4 replicates for the LGAs in Abuja FCT). The NBS did a listing of the households for all the EAs in the LGA master sample.

Given the variability in size of EAs, many countries select the EAs with probability proportional to size (PPS) at the first sampling stage, where the measure of size is based on the number of households from the census frame. One reason why NBS selected the EAs with equal probability within each LGA is that they were not provided with information on the number of households in the EAs from the 2006 Census data.

A2.2 NISH2 Master Sample

In examining the probabilities of selection for the NISH master sample by LGA within each state, it was found that the corresponding weights for each LGA vary by a factor of N_{SL}/n_{SL} (number of EAs in the 2006 Census frame for the LGA divided by the number of EAs selected in the LGA for the LGA master sample); that is, the sample households in the larger LGAs will have a higher weight. In order to reduce this variability in the weights for the sample households within each state, it is recommended to select the alternative NISH2 master sample with a more proportional allocation of the sample EAs by LGA within each state. For the NISH2, the sample of EAs within each state can be selected with probability proportional to the ratio N_{SL}/n_{SL} for each LGA. Since the NISH2 sample EAs are selected as a subsample of the LGA master sample EAs, the probabilities of selection are based on the corresponding probabilities for the LGA master sample multiplied by the subsampling rate for the NISH2 sample. This sampling procedure for the NISH2 master sample will result in an approximately proportional allocation of the sample EAs to the LGAs within each state. In order to illustrate the effect of this sampling procedure on the basic weights by state, we can examine the resulting probability of selection for the households in a survey based on the NISH2 sample, expressed as follows:

$$p_{SLi} = \frac{n_{SL}}{N_{SL}} \times \frac{n_S \times \left(\dfrac{N_{SL}}{n_{SL}} \right)}{\displaystyle\sum_{L \& S} \sum_{i \& L} \frac{N_{SL}}{n_{SL}}} \times \frac{m_{SLi}}{M'_{SLi}} = \frac{n_{SL}}{N_{SL}} \times \frac{n_S \times \left(\dfrac{N_{SL}}{n_{SL}} \right)}{\displaystyle\sum_{L \& S} n_{SL} \times \left(\dfrac{N_{SL}}{n_{SL}} \right)} \times \frac{m_{SLi}}{M'_{SLi}} =$$

$$= \frac{n_{SL}}{N_{SL}} \times \frac{n_S \times \left(\dfrac{N_{SL}}{n_{SL}}\right)}{\sum\limits_{L \varepsilon S} N_{SL}} \times \frac{m_{SLi}}{M'_{SLi}} = \frac{n_{SL}}{N_{SL}} \times \frac{n_S \times N_{SL}}{n_{SL} \times N_S} \times \frac{m_{SLi}}{M'_{SLi}} = \frac{n_S}{N_S} \times \frac{m_{SLi}}{M'_{SLi}},$$

where:

$p_{SLi} =$ overall sampling probability for households selected based on the NISH2 master sample in the i-th sample EA in LGA L within state S

$n_{SL} =$ number of sample EAs in LGA L of state S for the LGA master sample

$N_{SL} =$ total number of EAs in the 2006 Census frame for LGA L of state S

$n_S =$ number of EAs selected in subsample for NISH2 in state S

$m_{SLi} =$ total number of sample households selected in the i-th sample EA in LGA L of state S

$M'_{SLi} =$ number of households listed in the i-th sample EA in LGA L of state S

$N_S =$ total number of EAs in the 2006 Census frame for state S

The number of sample EAs in the LGA master sample for each state (nSL) is generally equal to 30, except for Abuja FCT, where 40 sample EAs were selected per LGA. However, some of the original sample EAs were found to be empty, so the actual value of nSL for some states is less than 30 (or less than 40 in the case of Abuja).

It can be seen in the simplified formula for the probability of selection of the sample households specified above that it is the same as if the sample EAs for each state had been selected with equal probability from all the census EAs in that state, regardless of the LGA. In this case the large LGAs would have a greater number of sample EAs in NISH2 than the small LGAs. The implicit allocation of the sample EAs by LGA would be approximately proportional to the total number of EAs in the LGA. The weights will still vary by the number of households listed in each sample EA.

During the selection of the sample EAs for NISH2, it was found that in some smaller states the larger LGAs had a measure of size N_{SL}/n_{SL} that was larger than the sampling interval, given the relatively small number of pooled EAs in the LGA master sample for the state. For example, in Bayelsa and Abuja FCT there are only 240 sample EAs each in the LGA master sample, so most of these EAs were selected in the NISH2 master sample of 200 EAs for each state. In these cases, all the 30 (or 40) sample EAs in the LGA master sample for these larger LGAs were selected with a probability of 1 for the NISH2 master sample. The remaining NISH2 sample EAs for each state were selected from the other LGAs with PPS, using the measure of size N_{SL}/n_{SL}.

A2.3   Sample Size and Allocation for the 2018/19 NLSS

The sample size for household surveys such as the NLSS is determined by the accuracy required for the survey estimates for each domain, as well as by the logistical, timing and resource constraints. The accuracy of the survey results depends on both the sampling error, which can be measured through variance estimation, and the nonsampling error, which results from all other sources of error, including response and measurement errors as well as coding, keying and processing errors. The sampling error is inversely proportional to the square root of the sample size. On the other hand, the nonsampling error may increase with the sample size, since it is more difficult to control the quality of a larger survey operation. It is therefore important that the overall sample size be manageable for quality and operational control purposes. This is especially important given the challenge of collecting accurate information on household income and expenditures, as well as crop area and production.

A stratified multi-stage sample design is used for the 2018/19 NLSS. The frame for the sample, the NISH2, is stratified by the 36 individual states of Nigeria and Abuja FCT. Although the master sample is not stratified by urban and rural areas, the individual EAs in the frame are classified by urban and rural areas, so it will be possible to obtain urban and rural estimates from the NLSS data at the national level that are consistent with the corresponding results from the previous NLSS. Since the NISH2 master sample is based on the same LGA master sample, the criteria for classifying urban and rural EAs are the same. The 200 EAs selected for the NISH2 within each state are implicitly allocated to the urban and rural strata approximately in proportion the number of urban and rural EAs in the frame for the state.

In the case of the 2009/10 HNLSS, the consumption module used for estimating the poverty rate was only administered to a sample of 5 of the 10 households selected in each sample EA. However, for the 2018/19 NLSS it would be more cost-effective to collect the consumption data for the full sample of 10 households in each sample EA. The design effect of the survey estimates depends on the number of households selected per EA, and it has been found in the Living Standards Measurement Surveys and similar household socioeconomic surveys with a consumption module in different countries that a sample of 10 households per cluster is approximately optimum. A smaller number of sample households per cluster would require a larger number of sample clusters and would increase the cost of the survey related to listing and transportation.

In order to examine the sample size for the 2018/19 NLSS, the estimates of standard errors, 95% confidence intervals and design effects from the 2009/10 HNLSS for estimates of the poverty rate at the state level shown in Annex A were used in a simulation study to estimate the approximate level of precision that can be expected for the 2018/19 NLSS based on different sample sizes. Given that 5 households were selected per sample EA for the consumption module of the 2009/10 HNLSS while 10 households will be selected per EA for 2018/19 NLSS, the corresponding design effect will be increased for the new survey. The design effect due to clustering in the 2009/10 HNLSS sample can be expressed as follows:

$$DEFF_{2009} = 1 + \delta_x \times \left(\bar{n}_{2009} - 1 \right),$$

where:

$DEFF_{2009} =$ tabulated design effect for estimate of the poverty rate based on NLSS 2009/10 sample design

$\delta_x =$ intraclass correlation coefficient (similarity of households within EA) for the poverty variable

$\bar{n}_{2009} =$ average number of households selected per cluster (EA) for the consumption module in the 2009/10 HNLSS (that is, 5 households)

It can be seen that the design effect depends on the number of households selected in each EA, as well as the correlation of households within the EA. Ignoring the components of the design effect related to stratification and differential weights, we can use the design effects from the 2009/10 HNLSS data to estimate the intraclass correlation coefficient as follows:

$$\delta_x = \frac{DEFF_{2009} - 1}{\bar{n}_{2009} - 1}$$

Assuming that this intraclass correlation coefficient has not changed over time, this estimated was used to calculate the approximate design effect based on the 2018/19 NLSS sample design as follows:

$$DEFF_{2018} = 1 + (\bar{n}_{2018} - 1) \times \frac{DEFF_{2009} - 1}{\bar{n}_{2018} - 1},$$

where:

$DEFF_{2018}$ = estimated design effect based on the sample design for 2018/19 NLSS

$\bar{n}_{2018}$ = average number of households selected per EA for 2018/19 NLSS (that is, 10)

This formula was used to estimate the increased design effects based on the 2018/19 NLSS sample design. The ratio of the variance (square of the standard error) for the survey estimate based on the sample design for 2018/19 NLSS to the corresponding variance based on the 2009/10 HNLSS data can be expressed as follows:

$$\frac{Var_{2018}(\hat{p})}{Var_{2009}(\hat{p})} = \frac{\frac{\hat{p} \times (1-\hat{p})}{n_{2018}} \times DEFF_{2018}}{\frac{\hat{p} \times (1-\hat{p})}{n_{2009}} \times DEFF_{2009}} \approx \frac{n_{2009}}{n_{2018}} \times \frac{DEFF_{2018}}{DEFF_{2009}},$$

where:

$Var_{2018}(\hat{p}) =$ approximate variance (square of standard error) for estimate of poverty rate based on the sample design for the 2018/19 NLSS

$Var_{2009}(\hat{p}) =$ variance for estimate of the poverty rate calculated from the NLSS 2009/10 data based on the actual sample design

$\hat{p} =$ poverty rate estimated from the 2009/10 HNLSS data

$n_{2018} =$ number of sample households (for domain of estimation) for 2018/19 NLSS

$n_{2009} =$ actual number of sample households (for domain of estimation) in NLSS 2009/10

Assuming that the poverty rate has not changed much since the 2009/10 HNLSS, this ratio of the variances simplifies as shown above. In this case, this ratio can be used for calculating the approximate standard error of the estimate that would result from the proposed sample design for 2018/19 NLSS as follows:

$$se_{2018}(\hat{p}) = \sqrt{Var_{2018}(\hat{p})} \approx \sqrt{Var_{2009}(\hat{p}) \times \frac{n_{2009}}{n_{2018}} \times \frac{DEFF_{2018}}{DEFF_{2009}}}$$

This formula was used with the 2009/10 HNLSS estimates of standard errors and design effects in order to determine the expected level of precision for the estimates of the poverty rate by state for the 2018/19 NLSS based on three alternative sample sizes per state: 500, 600 and 800 households. The full results of these calculations are provided upon request.

Examination of these three scenarios indicated that there is a relatively small decrease in the sampling errors and margins of error as the sample size increases. For the sampling alternative of 600 sample households per state, the margins of error for the estimates of the poverty rate were within 10% for more than half the states. In the case of the margins of error that exceed 10%, most were less than 12%. The states with a higher margin of error for the poverty rate generally have relatively high design effects, which should actually be lower for the 2018/19 NLSS.

Based on considerations of the survey budget and data quality, for the 2018/19 NLSS it was decided to select 60 EAs per state, with 10 sample households per sample EA, for a total of 600 sample households per state. The overall sample size at the national level is 2,220 sample EAs and 22,200 households.

A2.4   Sample Selection Procedures for the 2018/19 NLSS

Since the sample EAs for the 2018/19 NLSS were selected directly from the NISH2 master sample, the first sampling stages involved the selection process described previously for the LGA master sample and the NISH2 master sample. Since the NISH2 master sample for each state is composed of state-level replicates of 10 sample EAs each, a total of 6 replicates were selected from the NISH2 master sample for each state, to provide a total sample of 60 EAs per state. The 20 replicates of EAs in the NISH2 master sample for each state are systematic subsamples. Therefore the 6 replicates selected for the 2018/19 NLSS in each state were also selected using random systematic sampling. In this way the following NISH2 replicates were selected for the NLSS in each state: 2, 5, 8, 12, 15 and 18. This sampling procedure provides a similar distribution of the sample EAs within each state as if one systematic sample of 60 EAs had been selected directly. The implicit allocation of the 60 sample EAs by LGA within each state is approximately proportional to the total number of EAs in the 2006 Census frame for the LGA.

A2.5 Weighting Procedures for the 2018/19 NLSS

In order for the sample estimates from the 2018/19 NLSS data to be representative of the population, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample household is equal to the inverse of its probability of selection (calculated by multiplying the probabilities at each sampling stage). As described in the section on the NISH2 master sample, the probability of selection for the 2018/19 NLSS sample households in most states can be defined as follows:

$$p_{SLi} = \frac{n_{SL}}{N_{SL}} \times \frac{200 \times \left(\dfrac{N_{SL}}{n_{SL}}\right)}{\displaystyle\sum_{L \varepsilon S} N_{SL}} \times \frac{6}{20} \times \frac{10}{M'_{SLi}} = \frac{60}{N_S} \times \frac{10}{M'_{SLi}}$$

The first component of this probability corresponds to the LGA master sample and second component corresponds to the NISH2 master sample. As described previously, in the case of a few larger LGAs in small states such as Bayelsa and Abuja FCT, for the NISH2 master sample it was necessary to select all the sample EAs in the LGA master sample (generally 30 or 40 EAs). In these cases, the second component of this probability would be equal to 1.

The basic weight for the 2018/19 NLSS sample households is the inverse of this probability of selection, calculated as follows:

$$W_{SLi} = \frac{1}{p_{SLi}}$$

where:

$W_{SLi} = $ basic weight for the 2018/19 NLSS sample households in the i-th sample EA in

LGA L of state S

It is also important to adjust the weights to take into account the non-interview households in each sample EA. Since the weights will be calculated at the level of the sample EA, it would be advantageous to adjust the weights at this level. The final weight (W'SLi) for the sample households in the i-th sample EA in state S of LGA L can be expressed as follows:

$$W'_{SLi} = W_{SLi} \times \frac{10}{m'_{SLi}} \, ,$$

where:

$m'_{SLi} =$ number of sample households with completed interviews in the i-th sample EA in state S of LGA L