THE FEDERAL REPUBLIC
OF NIGERIA

NATIONAL BUREAU OF STATISTICS

Basic Information Document

Nigeria
General Household Survey–Panel
2018/19

Original: December 4, 2019
Latest revision: October 1, 2021

ACRONYMS

BMGF	Bill and Melinda Gates Foundation
FDE	First Data Entry
EA	Enumeration Area
FCT, Abuja	Federal Capital Territory, Abuja
FMA&RD	Federal Ministry of Agriculture and Rural Development
GHS	General Household Survey
GHS-Panel	General Household Survey-Panel (panel subcomponent of GHS)
HNLSS	Harmonized National Living Standards Survey
LGA	Local Government Area
LSMS-ISA	Living Standards Measurement Study – Integrated Surveys on Agriculture
NASS	National Agricultural Sample Survey
NBS	National Bureau of Statistics
NFRA	National Food Reserve Agency
SDE	Second Data Entry
TOT	Training of Trainers
WB	World Bank

Table of Contents

1.0 Introduction

The purpose of the present document is to provide detailed information on the General Household Survey-Panel (GHS-Panel) fielded by the National Bureau of Statistics (NBS) in 2018-2019. This survey is the fourth wave of a panel survey of households. The GHS-Panel is the result of a partnership that the NBS has established with the Bill and Melinda Gates Foundation (BMGF) and the World Bank (WB). The ability to follow the same households over time makes the GHS-Panel a powerful tool for studying and understanding income generating activities and socio-economic outcomes in Nigeria. The GHS-Panel is the first panel survey to be carried out by NBS.

The original GHS-Panel sample was subcomponent of the larger General Household Survey (GHS) survey. The GHS is a cross-sectional survey of 22,200 households carried out periodically throughout the country and was last conducted in 2010. Under the work of the partnership, a full revision of the questionnaire was undertaken and at the same time a sub-sample of the GHS was randomly selected to form the sample of the GHS-Panel. The original GHS-Panel consists of 5,000 households of the GHS collecting additional data on agricultural activities, other household income activities, and household expenditure and consumption. As the focus of this panel component is to improve data from the agricultural sector and link this to other facets of household behaviour and characteristics, the GHS-Panel questionnaire drew heavily on the Harmonized National Living Standards Survey (HNLSS – a multi-topic household survey) and the National Agricultural Sample Survey (NASS – the key agricultural survey). The fourth wave of the GHS-Panel was carried out in two visits (post-planting visit in July - September 2018 and post-harvest visit in January - February 2019).

This GHS-Panel is part of a larger, regional project in Sub-Saharan Africa to improve agricultural statistics. Nigeria is one of the eight countries being supported by the WB, through funding from the BMGF, to strengthen the production of household-level data on agriculture. This regional project, the Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) has the over-arching objective of improving our understanding of agriculture in Sub-Saharan Africa – specifically, its role in household welfare and poverty reduction.

The present document is designed to provide an overview of Wave 4 of the GHS-Panel. Wave 4 consisted of two visits to each household: the post-planting visit occurred directly after the planting season to collect information on preparation of plots, inputs used, labour used for planting, and other issues related to the planting season for the agriculture questionnaire as well as administer a household and community questionnaire. The post-harvest visit occurred after the harvest season and collected information on crops harvested, labour used for cultivating and harvesting activities, and other issues related to the harvest cycle for the agriculture questionnaire. A household and community questionnaire were also administered in the post-harvest visit.

The Basic Information Document for the GHS-Panel 2010/2011 (Wave 1), 2012/2013 (Wave 2) and 2015/2016 (Wave 3) all contain additional background information and should be used in conjunction with this document.

2.0 The Survey Instruments

The GHS-Panel Wave 4 consists of three questionnaires for each of the two visits. The ***Household Questionnaire*** was administered to all households in the sample. The ***Agriculture Questionnaire*** was administered to all households engaged in agricultural activities such as crop farming, livestock rearing and other agricultural and related activities. 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]

GHS-Panel Household Questionnaire*:* The Household Questionnaire provides information on demographics; education; health (including anthropometric measurement for children); labour; food and non-food expenditure; household nonfarm income-generating activities; food security and shocks; safety nets; housing conditions; assets; information and communication technology; and other sources of household income. Household location is geo-referenced in order to be able to later link the GHS-Panel data to other available geographic data sets (forthcoming).

GHS-Panel Agriculture Questionnaire*:* The Agriculture Questionnaire solicits information on land ownership and use; farm labour; inputs use; GPS land area measurement and coordinates of household plots; agricultural capital; irrigation; crop harvest and utilization; animal holdings and costs; and household fishing activities. Some information is collected at the crop level to allow for detailed analysis for individual crops.

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

The Household Questionnaire is slightly different for the two visits. Some information was collected only in the post-planting visit, some only in the post-harvest visit, and some in both visits. See Section 7.2.1 for more details.

The Agriculture Questionnaire collects different information during each visit, but for the same plots and crops. See Section 7.2.2 for more details.

[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).

The Community Questionnaire collected prices during both visits, and different community level information during the two visits. See Section 0 for more details.

The contents of each questionnaire for the GHS-Panel post-planting and GHS-Panel post-harvest are outlined below.

Table 2-1: GHS-Panel Household Wave 4 Questionnaire – Post-Planting Visit

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
3A	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)
3B	Time Use	Individuals 5 years and above	Domestic work at home
4A	Savings and Insurance	Individuals 15 years and above	Savings made and insurance
4B	ICT – Mobile Phone Banking	Individuals 10 years and above	Access to mobile phone and Internet
4C	Credit	Household head or most knowledgeable person	Credit history including loans received, loans pending, or loan refusals
5	Household Assets	Household head or most knowledgeable person	Ownership of assets and value
7A	Meals Away from Home	Most knowledgeable person	Value of food consumed outside the home during the last 7 days
7B	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

Section	Topic	Respondent	Description
8	Non-food Expenditure	Person responsible for household purchases	Non-food expenditure during the last week/last month/last 6 months/last 12 months
9	Food Security	Person responsible for food preparations or food purchases	Food security status of the household during the last 7 days/last 12 months
11	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)

Table 2-2: GHS-Panel Agriculture Wave 4 Questionnaire – Post-Planting Visit

Section	Topic	Respondent	Description
Cover	Cover	Field staff	Household identifiers, enumerator and supervisor identifiers, date of interview
11A	Plot Roster	Farmer, owner or manager of plot	Roster of all plots owned and/or managed by the household, farmer estimated area, GPS measured area, GPS measured location of the plot, decision maker of the plot
11B1	Land Inventory	Most knowledgeable person on household farm operations	Plot acquisition, tenure (legal title, ownership document), rental arrangement, sharecropping arrangement, use rights (collateral, inheritance)
11C1	Household Labour Hired and Exchange Labour	Most knowledgeable person on household farm operations	Household labour that worked on the plot. Includes information on amount of time spent by each person Hired and free/exchange labour that worked on the plot. Includes information on amount of time spent by each person-type and payments made to hired labour
11E	Seed Acquisition	Most knowledgeable person on household farm operations	Source (left over, free and purchased), quantity and cost of seeds used (including transportation cost)
11F	Planted Crops	Most knowledgeable person on household farm operations	Roster of all the field and tree crops cultivated on the plot, seeds planted, last completed production of tree/permanent crops, expected harvest
11I	Animal Holdings	Owner or caretaker of animals	Farm animals owned by the household and commercial activity with these animals

Section	Topic	Respondent	Description
11J	Animal Costs	Owner or caretaker of animals	Livestock farmer caretaker activities and costs
11K1	Animal Power and Dung	Owner or caretaker of animals	Earnings from animal power and animal dung
11K2	Milk Production	Owner or caretaker of animals	Amount of production and commercial activity with milk
11K3	Egg Production	Owner or caretaker of animals	Amount of production and commercial activity with eggs
11L1	Extension Services	Most knowledgeable person	Main source (government and non-governmental) of agricultural advice
11L2	Extension Sources	Most knowledgeable person	Services provided by main source of agricultural advice. Includes frequency of visits and usefulness of the information provided

Table 2-3: GHS-Panel Community Wave 4 Questionnaire – Post Planting Visit

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 and education level
C2	Food Prices	Market Food Sellers	Prices of food items in the community
C3	Labour	Community Focus Group	Labour hired for agricultural activities by famers in the community
C4	Land Prices and Credit	Community Focus Group	Land acquisition, land prices and credit

Table 2-4: GHS-Panel Household Wave 4 Questionnaire – Post Harvest Visit

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	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)
4	Health	All individuals	General health status, health care utilization and cost of health services, functioning and disability status, anthropometrics and use of bed net
6	Remittances	Individuals 10 years and above	Remittances received (from abroad and from within Nigeria)
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
10A	Meals Away from Home	Most knowledgeable person	Value of food consumed outside the home during the last 7 days
10B	Food Expenditures	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
10C	Aggregate Food Consumption	Person responsible for food preparations or food purchases	Number of days the household consumed each food group item during the last 7 days
11	Non-food Expenditures	Person responsible for household purchases	Non-food expenditure during the last week/last month/last 6 months/last 12 months

Section	Topic	Respondent	Description
12	Food Security	Person responsible for food preparations or food purchases	Food security status of the household during the last 7 days/last 12 months
13	Other Household Income	Household head or most knowledgeable person	Miscellaneous income received by the household
14	Safety Nets	Household head or most knowledgeable person	Household access to and utilization of safety nets
15A	Economic Shocks	Household head or most knowledgeable person	Economic shocks affecting the household and coping mechanism adopted by the household

Table 2-5: GHS-Panel Agriculture Wave 4 Questionnaire – Post Harvest Visit

Section	Topic	Respondent	Description
Cover	Cover	Field staff	Household identifiers, enumerator and supervisor identifiers, date of interview
A1	Plot Roster	Farmer, owner or manager of plot	Follow-up on use of land listed in post-planting visit, farmer estimated area, GPS measured area, GPS measured location of the plot, decision maker of the new plots (i.e. added since post-planting visit)
A2a	Household Labour	Most knowledgeable person on household farm operations	Household labour that worked on the plot. Includes information on amount of time spent by each person
A2b	Hired and Exchange Labour	Most knowledgeable person on household farm operations	Hired and free/exchange labour that worked on the plot. Includes information on amount of time spent by each person-type and payments made to hired labour
11C2	Input Use on Plot	Farmer, owner or manager of plot	Pesticide use, herbicide use, inorganic fertilizer use, organic fertilizer use, animal traction, equipment and machinery use
11C3	Input Acquisition	Most knowledgeable person on household farm operations	Cost of purchased input, organic fertilizer from own production, cost of machinery
A3i	Field Crop Harvest	Most knowledgeable person on household farm operations	Timing of harvest, quantity of harvested crops, estimated value, expected harvest
A3ii	Crop Disposition	Most knowledgeable person on household farm operations	Sale of unprocessed crops, processing of crops, sale of processed crops, other crop disposition

A3iii	Tree and Permanent Crop Harvest	Most knowledgeable person on household farm operations	Timing of harvest, quantity of harvested crops, estimated value
PHL1	Post-Harvest Labour - Household	Most knowledgeable person on household post-harvest operations	Household labour that worked on post-harvest activities of the harvested crops. Includes information on amount of time spent by each person
PHL2	Post-Harvest Labour – Hired and Exchange	Most knowledgeable person on household post-harvest operations	Hired and free/exchange labour that worked on post-harvest activities of the harvested crops. Includes information on amount of time spent by each person-type and payments made to hired labour
A4	Agricultural Capital	Household head or most knowledgeable person	Ownership and value of agricultural machinery and tools owned by the household
A5b	Extension Services	Most knowledgeable person	Services provided by various source of agricultural advice (government and non-government). Includes frequency of visits and usefulness of the information provided
A9a	Fishing	Household member(s) responsible for fishing activities	Capture/ harvesting, sales and consumption of captured/harvested fish

Table 2-6: GHS-Panel Community Wave 4 Questionnaire – Post-Harvest Visit

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 and education level
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	Community Resource Management	Community Focus Group	Resource availability and the resource management in the community

C5	Community Changes	Community Focus Group	Changes in the community compared to 3 years ago
C6	Community Key Events	Community Focus Group	Key events in the community in the last 3 years
C7	Community Needs, Actions, and Achievements	Community Focus Group	Discussion on the needs in the community
C8	Food Prices	Market Food Sellers	Prices of food items in the community
C9	Labour	Community Focus Group	Labour hired for agricultural activities by male and female farmers in the community

There were some changes made in the questionnaires between Waves 3 and 4 to improve the questionnaire while still maintaining comparability between the two waves as much as possible. When questions were dropped or added, every effort was made to keep question numbers consistent with previous waves. If new questions were added in the middle of a section, letters were added to the question number (for example a new question added between Q21 and Q22 would be Q21a). Table 2-7 and Table 2-8 outline these changes for the post-planting and post-harvest visits, consecutively.

Table 2-7: Wave 3 to Wave 4 Comparison, Post-Planting

Questionnaire	Section	Notes
Household Questionnaire	Cover	Questions dropped in Wave 4: AG1, AG2, AG3
		Questions added in Wave 4: AG1a, AG2a, AG4, AG5
	Section 1: Household Roster	Questions added in Wave 4: Q4a, Q4b, Q8a, Q13a1, Q13b1
		Questions dropped in Wave 4: Q9-Q10, Q13b
		Questions re-worded or options modified in Wave 4: Q8, Q22, Q23, Q27, Q28, Q29
		Q11 in Wave 3 was integrated to Q13 in Wave 4
		In Wave 4, Q13 was asked for up to 6 spouses
		In Wave 4, Q18 was asked to the household head and spouse only
		Q32 in Wave 3 was split into Q32b and Q32a in Wave 4
	Section 3: Labour	Questions added in Wave 4: Q7a-c, Q8a-e, Q9a-b, Q11a, Q11b, Q15c1, Q15g-k, Q38b
		Questions dropped in Wave 4: Q6c, Q8, Q9, Q25-Q31, Q34-Q37, Q38a-Q40
		Questions re-worded or options modified in Wave 4: Q10, Q11, Q12b, Q15c, Q23, Q38
		Q21 in Wave 3 was split into Q21 and Q21b in Wave 4
		Q22 modified to “Select all that apply” in Wave 4
		Q24 was re-worded and split into Q24a and Q24b in Wave 4

Questionnaire	Section	Notes
	Section 3B: Time Use	Added in Wave 4
	Section 4A: Savings and Insurance	Questions added in Wave 4: Q3a, Q17b, Q17c
		Option "Don't know" added in Wave 4: Q3-Q7
		Questions re-worded or options modified in Wave 4: Q2
		Q9 modified to "Select all that apply" and options modified in Wave 4
		In Wave 4, Q16 is asked at household level
		In Wave 4, Q17a is asked at household level, modified to "Select all that apply", and dropped insurance companies codes
	Section 4B: ICT – Mobile Phone Banking	Questions dropped in Wave 4: Q10-Q10h, Q15a
		Questions re-worded or options modified in Wave 4: Q9, Q15
	Section 4C: Credit	Questions added in Wave 4: Q18-Q27,
		Questions dropped in Wave 4: Q3, Q5, Q6, Q12-Q15
		Questions re-worded or options modified in Wave 4: Q1, Q2, Q4, Q17
	Section 5: Household Assets	Change in Wave 4: Changes to item codes
		Questions added in Wave 4: Q2a
	Section 7A: Meals Away from Home	No changes
	Section 7B: Food Expenditure	Change in Wave 4: Changes to food item and unit codes
		Questions re-worded or options modified in Wave 4: Q2, Q5-Q7
		Questions added in Wave 4: Q3a, Q8
		Q3-Q4 became Q9-Q10 and are asked on "past 30 days" in Wave 4
		Q5-Q7 no longer have unit codes in Wave 4
	Section 8: Non-food Expenditure	Change in Wave 4: Changes to non-food item codes
	Section 9: Food Security	Questions added in Wave 4: Q8
		Questions dropped in Wave 4: Q1, Q3, Q4
		Questions re-worded or options modified in Wave 4: Q7
	Section 9B: Subjective Wellbeing	Dropped in Wave 4
	Section 11: Housing	Questions added in Wave 4: Q5a-Q5c, Q40-Q75
		Questions dropped in Wave 4: Q10-Q28i, Q31-32, Q34-Q35, Q37

Questionnaire	Section	Notes
		Questions re-worded or options modified in Wave 4: Q1, Q3, Q5-Q8, Q33b, Q36, Q38
		In Wave 4, Q33 (a) Dry season dropped
Agriculture Questionnaire	Section 11A: Plot Roster	In Wave 4, Q27 (in Section 11B) is asked after Q5 in Section 11A
		Questions re-worded or options modified in Wave 4: Q6
	Section 11B1: Land Inventory	Questions added in Wave 4: Q8b1-Q8b4, Q10a-Q10e, Q11, Q12, Q13a, Q14b, Q16a, Q14a1, Q14b1, Q23a-Q23c, Q25a, Q26a, Q29a, Q35a, Q35b, Q37a, Q37b
		Questions dropped in Wave 4: Q11, Q12, Q15, Q23, Q29, Q30, Q36, Q38
		Questions re-worded or options modified in Wave 4: Q4, Q8a, Q13, Q14, Q16, Q25, Q28, Q28a
		In Wave 4, Q17 is asked before Q16a and Q16
		Questions modified to “Select all that apply” in Wave 4: Q22, Q41, Q50
		In Wave 4, population of Q49 is modified
		In Wave 4, Q11 and Q12 from Section 11C1 moved to Section 11B1 and became Q11a and Q11b
	Section 11C1: Household Labor, Hired and Exchange Labor	Questions dropped in Wave 4: Q1
		Questions added in Wave 4: Q1a-Q1d, Q2a-Q10a, Q13-Q17
		Q11 and Q12 moved to Section 11B1
	Section 11E: Seed Acquisition	In Wave 4, all the questions are asked at crop level and no longer use the plot-crop roster.
		Questions dropped in Wave 4: Q3-3b, Q5, Q7, Q9, Q12, Q16, Q17, Q22-Q37
		Questions added in Wave 4: Q4a, Q21a, Q22a, Q23a
		In Wave 4, Q12 and Q19 are asked for all the crops and not per crop.
	Section 11F: Planted Field and Tree Crops	Section renamed “Planted Crops” in Wave 4
		Questions dropped in Wave 4: Q2, Q5, Q14, Q15
		Questions added in Wave 4: Q2a, Q3aa, Q3b-Q3f, Q4aa, Q5aa, Q8aa
	Section 11I: Animal Holdings	Change in Wave 4: Changes to animal codes
		Questions dropped in Wave 4: Q1, Q19-Q24
		Questions added in Wave 4: Q1a, Q2a, Q2b, Q6a-Q6d, Q19a-c, Q6p-Q19cp
		Questions re-worded or options modified in Wave 4: Q2, Q12, Q18

Questionnaire	Section	Notes
		Questions modified to “Select all that apply” in Wave 4: Q4, Q5
		In Wave 4, Q3 is asked after Q5
		In Wave 4, Q6 changed to “12 months ago”
		In Wave 4, Q7-Q18 changed to “past 12 months”
	Section 11J: Animal Costs	In Wave 4, questions are asked at animal type. All questions are new
	Section 11K1: Animal Power and Dung	Added in Wave 4
	Section 11K: Agricultural By-product	Split into 2 sections (Section 11K2: Milk Production and Section 11K3: Egg Production) in Wave 4
	Section 11K2: Milk Production	New section in Wave 4, split from Section 11K in Wave 3
		In Wave 4, all the questions are asked at animal type. All questions are new
	Section 11K3: Egg Production	New section in Wave 4, split from Section 11K in Wave 3
		In Wave 4, all the questions are asked at animal type. All questions are new
	Section 11L1: Extension Services I	Section renamed “Extension services” in Wave 4
	Section 11L1: Extension Services II	Section renamed “Extension sources” in Wave 4
		Questions modified to “Select all that apply” in Wave 4: Q2
		Option “Don’t know” added in Wave 4: Q9
	Section 12: Network Roster	Dropped in Wave 4
Community Questionnaire	Section C1: Respondents Characteristics	Questions re-worded or options modified in Wave 4: Q6
	Section C2a: Food Prices – Location 1	Section became “Section C2: Food Prices” in Wave 4
	Section C2b: Food Prices – Location 2	Dropped in Wave 4
	Section C2: Food Prices	New section in Wave 4: renamed from Section C2a in Wave 3
		Only one location in Wave 4
		Change in Wave 4: Changes to food item and unit codes
	Section C3: Labour	Questions added in Wave 4: Q2b, Q11a, Q12a
		Change in Wave 4: Changes to crop codes
	Section C4: Land Prices and Credit	Questions added in Wave 4: Q4a-d
		Questions dropped in Wave 4: Q4, Q8, Q11, Q14

Table 2-8: Wave 3 to Wave 4 Comparison, Post-Harvest

Questionnaire	Section	Notes
Household Questionnaire	Cover	Questions added in Wave 4: AG1a, AG2a, AG4
		Questions dropped in Wave 4: AG1, AG2, AG3
	Section 1: Household Roster	Questions added in Wave 4: Q4b-c, Q5a, Q5b, Q8a-c, Q13a-c, Q31c
		Questions dropped in Wave 4: Q9, Q12a-c, Q34-Q41
		In Wave 4, Q12 was asked for up to 6 spouses
		In Wave 4, Q17 was asked to the household head and spouse only
		Questions re-worded or options modified in Wave 4: Q21, Q22, Q26, Q27
		Q31 in Wave 3 was split into Q31b and Q31a in Wave 4
	Section 2: Education	Questions are for all household members 3 years and above in Wave 4
		Questions added in Wave 4: Q13a-c
		Questions re-worded or options modified in Wave 4: Q7, Q9, Q10, Q14, Q15, Q17, Q18
		In Wave 4, Q23 was split into Q23a and Q23b with new expenditure items
	Section 3: Labour	Questions added in Wave 4: Q7a-d, Q8a, Q8b, Q8b1, Q8b2, Q8c-g, Q9a, Q10a-g, Q15c1, Q50, Q51
		Questions dropped in Wave 4: Q6c, Q8-Q11, Q15f, Q25-Q31, Q35-Q49
		Questions re-worded or options modified in Wave 4: Q5, Q8, Q23, Q24
		Questions modified to “Select all that apply” in Wave 4: Q22
		In Wave 4, Q7 includes additional condition “Q7a”
		Q12b renamed Q12b1 in Wave 4
		Q15b renamed Q15b1 in Wave 4
		Q15dk is added in Wave 4 and Q15d and Q15f were consolidated to Q15dk
		Option “Don’t know” added in Wave 4: Q15c, Q15e
		Q16 modified and became Q16a in Wave 4
	Section 4A: Health	Questions added in Wave 4: Q25-Q33, Q54a-b, Q55-Q60
		Questions dropped in Wave 4: Q18-Q21, Q54
		Questions re-worded or options modified in Wave 4: Q2, Q3b, Q6
		In Wave 4, Q3 is asked to a different population due to new options and skips in Q2
		Q7 and Q8 are asked for only the most important consultation selected in Q6

Questionnaire	Section	Notes
		Q10 has modified instruction for interviewers in Wave 4
	Section 4B: Child Development	Dropped in Wave 4
	Section 6: Remittances	Questions added in Wave 4: Q0a-b, Q4c-d, Q11-Q17
		Questions re-worded or options modified in Wave 4: Q1, Q2, Q5, Q8, Q10
		Questions modified to "Select all that apply" in Wave 4: Q6, Q9
	Section 6A: Behavior	Dropped in Wave 4
	Section 6B: Attitude	Dropped in Wave 4
	Section 9: Nonfarm Enterprises and Income Generating Activities	Questions added in Wave 4: Q0a-b, Q1c-d, Q4c, Q12a, Q16a, Q29a, Q30a
		Questions dropped in Wave 4: Q7, Q8
		Questions re-worded or options modified in Wave 4: Q18, Q28
		In Wave 4, Q13 and Q13a were consolidated to Q13
	Section 10A: Meals Away from Home	No changes
	Section 10B: Food Expenditure	Change in Wave 4: Changes to food item and unit codes
		Questions added in Wave 4: Q3a
		Questions re-worded or options modified in Wave 4: Q2, Q5-Q7
		Q3-Q4 became Q9-Q10 and are asked on "past 30 days" in Wave 4
		Q5-Q7 no longer have unit codes in Wave 4
	Section 10C: Aggregate Food Consumption	Questions dropped in Wave 4: Q9-Q11
	Section 11: Non-food expenditure	Change in Wave 4: Changes to non-food item codes
		Questions dropped in Wave 4: Q9-Q11
	Section 12: Food Security	Questions added in Wave 4: Q8, Q2b1,
		Questions dropped in Wave 4: Q1, Q3, Q4
		Questions re-worded or options modified in Wave 4: Q7
	Section 13: Other Household Income	Re-organized to use a roster of income sources in Wave 4. All the questions are asked at income-source level
		Questions dropped in Wave 4: Q4
	Section 14: Safety Nets	Change in Wave 4: Changes to programme codes
		Questions added in Wave 4: Q1a-c, Q4a, Q6
		In Wave 4, Q1 is not asked to all households due to new questions and skips (Q1a-b)

Questionnaire	Section	Notes
	Section 15A: Economic Shocks	Questions added in Wave 4: Q7
		Questions dropped in Wave 4: Q5, Q6
	Section 15B: Deaths	Dropped in Wave 4
	Section 15C: Conflicts	Dropped in Wave 4
Agriculture Questionnaire	Section A1: Land	Section renamed “Plot Roster” in Wave 4
	Section A1: Plot Roster	New section in Wave 4: Renamed from Section A1 in Wave 4
		Questions added in Wave 4: Q1a, Q1c, Q1d, Q2b, Q2c, Q8a-Q8c
		Questions dropped in Wave 4: Q3, Q4b, Q8-Q25
	Section A2: Harvest Labour	Split into 2 sections (Section A2a: Household Labour and Section A2b: Hired and Exchange Labour) in Wave 4
	Section A2a: Household Labour	New section in Wave 4, split from Section A2 in Wave 3
		Section is entirely reformatted. Most information collected in Wave 3 are retained but collected in a different manner
		New information collected: Activities performed (Q1d)
	Section A2b: Hired and Exchange Labour	New section in Wave 4, split from Section A2 in Wave 3
		Section is entirely reformatted. Most information collected in Wave 3 are retained but collected in a different manner
		New information collected: Number of hours per day worked (Q3a, Q6a, Q9a), Activities performed (Q4a, Q7a, Q10a.)
	Section 11C2: Input Cost	Consolidated with Section 11D and split into 2 sections (Section 11C2: Input Use on Plot and Section 11C3: Input Acquisition) in Wave 4
	Section 11D: Fertilizer Acquisition	Consolidated with Section 11C and split into 2 sections (Section 11C2: Input Use on Plot and Section 11C3: Input Acquisition) in Wave 4
		Q1a, Q36 and Q37 moved to Section 11C2
		Questions dropped in Wave 4: Q1, Q2-Q35, Q38-Q41
	Section 11C2: Input Use on Plot	New section in Wave 4, split from Sections 11C2 and 11D in Wave 3
		Questions added in Wave 4: Q36-Q41
		Questions dropped in Wave 4: Q3-Q9, Q12-Q18, Q22, Q24, Q28, Q30-Q33
		In Wave 4, Q1a, Q36 and Q37 from Section 11D moved to Section 11C2
	Section 11C3: Input Acquisition	New section in Wave 4, split from Sections 11C2 and 11D in Wave 3
		All questions are new except for Q29, which moved from Section 11C2

Questionnaire	Section	Notes
	Section A3i: Agricultural Production – Harvest of Field and Tree Crops	Section renamed “Field Crop Harvest” in Wave 4
	Section A3i: Field Crop Harvest	New section in Wave 4: Section renamed from Section A3i in Wave 3
		Questions added in Wave 4: Q4b, Q4c
		Changes in Wave 4: Changes to crop and unit codes
	Section A3ii: Agricultural Production – Crop Disposition	Section renamed “Crop Disposition” in Wave 4
		Changes in Wave 4: Changes to crop and unit codes
		Questions re-worded or options modified in Wave 4: Q4
	Section A3ii: Crop Disposition	New section in Wave 4: Renamed from Section A3ii in Wave 3
		Questions added in Wave 4: Q10a, Q11aa, Q12a, Q13a, Q14a, Q15a, Q16a, Q17a, Q18a, Q20a
		Questions dropped in Wave 4: Q11-Q17, Q19, Q20, Q24
	Section A3iii: Tree and Permanent Crop Harvest	Added in Wave 4
	Section PHL1: Post-Harvest Labour - Household	Added in Wave 4
	Section PHL2: Post-Harvest Labour – Hired and Exchange	Added in Wave 4
	Section A4: Agricultural Capital	Change in Wave 4: Changes to item
		Questions added in Wave 4: Q1a, Q2a
		Questions dropped in Wave 4: Q3, Q3b, Q4, Q5, Q8
	Section A6: Animal Holdings	Dropped in Wave 4
	Section A7: Animal Cost	Dropped in Wave 4
	Section A5a: Extension Services (Topics)	Dropped in Wave 4
	Section A5b: Extension Services (Sources)	Section renamed “Extension Services” in Wave 4

Questionnaire	Section	Notes
	Section A5b: Extension Services	New section in Wave 4: renamed from Section A5b in Wave 3
		Questions added in Wave 4: Q1a, Q3a
		Questions dropped in Wave 4: Q4-Q8
	Section A8: Other Agricultural Income	Dropped in Wave 4
	Section A9a: Fishing	Questions are asked per source of fish and not per type of fish in Wave 4. All questions are new
	Section A9b: Fishing Capital and Revenues	Dropped in Wave 4
	Section A10: Network Roster	Dropped in Wave 4
Community Questionnaire	Section C1: Respondent Characteristics	Questions re-worded or options modified in Wave 4: Q6
	Section C2: Community Infrastructure and Transportation	No changes
	Section C3: Community Organizations	No changes
	Section C4: Community Resource Management	No changes
	Section C5: Community Changes	No changes
	Section C6: Community Key Events	In Wave 4, key events are pre-listed and description of events are no longer asked.
		Questions re-worded or options modified in Wave 4: Q1
	Section C6A: Conflict	Dropped in Wave 4
	Section 7: Community Needs, Actions and Achievements	Questions dropped in Wave 4: Q5-Q8
	Section C8a: Food Prices – First Location	Section became “Section C8: Food Prices” in Wave 4

Questionnaire	Section	Notes
	Section C8ba: Food Prices – First Location	Dropped in Wave 4
	Section C8: Food Prices	New section in Wave 4: renamed from Section C8a in Wave 3
		Only one location in Wave 4
		Change in Wave 4: Changes to food item and unit codes
	Section C9: Labour	Added in Wave 4

3.0 Wave 4 Sample and Weights

The original GHS-Panel sample of 5,000 households across 500 enumeration areas (EAs) was designed to be representative at the national level as well as at the zonal level. The complete sampling information for the GHS-Panel is described in the Basic Information Document for GHS-Panel 2010/2011. However, after a nearly a decade of visiting the same households, a partial refresh of the GHS-Panel sample was implemented in Wave 4. The refresh was conducted in order to maintain the integrity and representativeness of the sample. Overall attrition since the first wave was a modest 8.3 percent but was concentrated in some zones (19.5% in North East, 14% in South West). In addition to attrition, the refresh was also implemented to alleviate the burden on originally sampled households and to counter the potential for respondent conditioning. After being visited across 3 waves, respondents become familiar with the survey and this knowledge can alter their responses in a few ways: (1) being interviewed and the awareness that they will be interviewed again in the future may alter the behaviour of households (the so-called "Hawthorne effect") and (2) respondents learn the process of the interview and alter their responses to shorten the length of the interview (e.g. responding "No" to a gateway question means subsequent questions will not be asked). Both of these mechanisms could compromise both the accuracy and representativeness of the data.

For the partial refresh of the sample, a new set of 360 EAs were randomly selected which consisted of 60 EAs per zone. The refresh EAs were selected from the same sampling frame as the original GHS-Panel sample in 2010 (the "master frame"[2]). A listing of all households was conducted in the 360 EAs and 10 households were randomly selected in each EA, resulting in a total refresh sample of approximated 3,600 households.

In addition to these 3,600 refresh households, a subsample of the original 5,000 GHS-Panel households from 2010 were selected to be included in the new sample. This "long panel" sample was designed to be nationally representative to enable continued longitudinal analysis for the sample going back to 2010. The long panel sample consisted of 159 EAs systematically selected across the 6 geopolitical Zones. The systematic selection ensured that the distribution of EAs across the 6 Zones (and urban and rural areas within) is proportional to the original GHS-Panel sample. Interviewers attempted to interview all households that originally resided in the 159 EAs and were successfully interviewed in the previous visit in 2016. This includes households that had moved away from their original location in 2010. In all, interviewers attempted to interview 1,507 households from the original panel sample. If a household had moved from the location where they were found in the previous interview, the survey teams attempted to track and interview this household in its new location.

The combined sample of refresh and long panel EAs consisted of 519 EAs. This combined sample of households will be visited in subsequent waves of the GHS-Panel and therefore comprises a

[2] The master frame is a subsample of EAs from the 2006 census. See the Basic Information Document for the GHS-Panel Wave 1 for more information on the master frame.

new panel of households. While the combined sample generally maintains both national and Zonal representativeness of the original GHS-Panel sample, the security situation in the North East of Nigeria prevented full coverage of the Zone. Due to security concerns, rural areas of Borno state were fully excluded from the refresh sample and some inaccessible urban areas were also excluded. Security concerns also prevented interviewers from visiting some communities in other parts of the country where conflict events were occurring. Refresh EAs that could not be accessed[3] were replaced with another randomly selected EA in the Zone so as not to compromise the sample size. As a result, the combined sample is representative of areas of Nigeria that were accessible during 2018/19. The sample will not reflect conditions in areas that were undergoing conflict during that period. This compromise was necessary to ensure the safety of interviewers.

Table 3.1 Shows the detailed breakdown of the Wave 4 sample across the refresh and long panel samples. The sample shown in the table is the ultimate sample of households that was successfully interviewed in both Wave 4 visits. The final sample consisted of 4,976 households of which 1,425 were from the long panel sample and 3,551 from the refresh sample. Although 159 long panel and 360 refresh EAs were selected and visited in the post-planting visit, conflict events prevented interviewers from visiting 2 rural EAs in the North West during the post-harvest visit (one EA from the long panel sample and one from the refresh). Therefore, the final number of EAs visited in both post-planting and post-harvest was 158 long panel EAs and 359 refresh EAs.

Table 3.1: Final Sample Composition

Zone		Long Panel Sample		Refresh Sample		Combined Sample	
		# of EAs	#of HH	# of EAs	#of HH	# of EAs	#of HH
NORTH CENTRAL							
	Urban	7	61	18	176	25	237
	Rural	19	181	42	420	61	601
	Total	26	242	60	596	86	838
NORTH EAST							
	Urban	3	28	10	98	13	126
	Rural	21	200	50	500	71	700
	Total	24	228	60	598	84	826
NORTH WEST							
	Urban	5	46	12	120	17	166
	Rural	22	211	47	470	69	681
	Total	27	257	59	590	86	847
SOUTH EAST							
	Urban	7	61	15	146	22	207
	Rural	19	175	45	445	64	620

[3] Overall, 34 refresh EAs were inaccessible during the listing period or post-planting visit. The EAs were highly concentrated in the North East and North Central Zones where conflict (insurgency and farmer-herder attacks) were prevalent during this period.

	Total	26	236	60	591	86	827
SOUTH SOUTH							
	Urban	8	63	18	177	26	240
	Rural	18	158	42	416	60	574
	Total	26	221	60	593	86	814
SOUTH WEST							
	Urban	21	179	43	418	64	597
	Rural	8	62	17	165	25	227
	Total	29	241	60	583	89	824
TOTAL							
	Urban	51	438	116	1,135	167	1,573
	Rural	107	987	243	2,416	350	3,403
	TOTAL	**158**	**1,425**	**359**	**3,551**	**517**	**4,976**

3.1 Attrition in the Long Panel Sample

Since Wave 1, every effort has been made to track and interview households that had moved away from their original EA and keep attrition to a minimum. These efforts continued in Wave 4, particularly for the long panel sample. Household that had moved away from their previous location were interviewed in a separate tracking phase following the post-planting and post-harvest visits. Table 3.2 presents information on movement and attrition of long panel households in EAs retained for the Wave 4 sample. Of the 1,590 households interviewed in Wave 1 within these 159 EAs, 1,425 households were successfully interviewed in both visits of Wave 4. This implies an overall attrition rate since 2010 across these EAs of 10.4 percent. However, attrition is highly variable across Zones and sectors. The highest attrition was found in rural EAs in South West (22.5%) and lowest in rural EAs in North Central (4.7%). Attrition was also higher among urban (14.1%) than rural (8.6%) households. The Table further indicates the number of households that were interviewed in their original location and those that had moved and were interviewed in their new location. Overall, 152 long panel households had moved and were interviewed in their new location (over 10% of the sample). The higher number of households that had moved was in urban South West at 46 households (25.7% of the sample).

		Original Sample* (2010)	**Successfully Interviewed in W4 (2019)**			**Attrition (%)**
			Original Location	**Moved (Tracked)**	**Total**	
NORTH CENTRAL						
	Urban	70	50	11	61	12.9
	Rural	190	176	5	181	4.7
	Total	260	226	16	242	6.9
NORTH EAST						
	Urban	30	24	4	28	6.7

	Rural	210	195	5	200	4.8
	Total	240	219	9	228	5.0
NORTH WEST						
	Urban	50	42	4	46	8.0
	Rural	230	204	7	211	8.3
	Total	280	246	11	257	8.2
SOUTH EAST						
	Urban	70	56	5	61	12.9
	Rural	190	167	8	175	7.9
	Total	260	223	13	236	9.2
SOUTH SOUTH						
	Urban	80	46	17	63	21.3
	Rural	180	131	27	158	12.2
	Total	260	177	44	221	15.0
SOUTH WEST						
	Urban	210	133	46	179	14.8
	Rural	80	49	13	62	22.5
	Total	290	182	59	241	16.9
TOTAL						
	Urban	510	351	87	438	14.1
	Rural	1,080	922	65	987	8.6
	TOTAL	**1,590**	**1,273**	**152**	**1,425**	**10.4**

*Among the 159 EAs selected for the long panel.

3.2 Survey Weights

When a sample of households is selected for a survey, these households represent the entire population of the country. To accurately use the data to extrapolate results for the underlying population, weights must be applied that reflect the sampling strategy as well as the distribution of the underlying population.

In wave 4 of the GHS-Panel, there are two components of the sample: the refresh and the long panel samples. Two sets of weights were constructed for two different types of analysis. The first set of weights are those for the combined wave 4 sample. They can be used for cross-sectional analysis for the full GHS-Panel wave 4 sample (refresh plus long panel sample). The second set of weights are designed for longitudinal/panel analysis using the long panel sample only. These longitudinal weights can be used for analysis that seeks to track dynamics within long panel households across the 4 waves of the GHS-Panel. When calculating both weights, only households successfully interviewed in both visits of Wave 4 were considered.

The cross-sectional weights were constructed in three stages:

1. Base weights were calculated according to the inverse probability of selection for each household in the sample. In its simplest form, this weight reflects the two-stage design and thus is the product of the probability that the EA was selected from the frame and the probability that the household was selected within the EA.
2. The base weights were then adjusted for non-response within the EA (the ratio of households successfully interviewed and households selected).
3. The weights were calibrated to reflect the distribution of the underlying population. The weights were calibrated to (1) reflect the total number of households in each Zone in 2010 (i.e. during the first wave of the GHS-Panel) according to population predictions from the 2006 Census and (2) reflect the total number of persons as estimated in the (weighted) Wave 3 sample of the GHS-Panel. The calibration in (1) was performed to maintain consistency with the calibration methodology adopted in previous rounds of the GHS-panel.
4. Lastly, outlier weights were trimmed with a lower bound of 400 and an upper of 50,000 while also maintaining the calibrated parameters in step 3.

The long-panel weights were constructed in the following stages:

1. The original design weights (from wave 1) served as the base weights
2. The base weights were adjusted to account for selection of the EA into the long-panel sample. The adjustment was performed for with a simple ratio adjustment that maintains the sum of weights within each strata (Zone).
3. Another ratio adjustment was performed to account for household non-response and attrition between wave 1 visit 1 and wave 4 visit 2.
4. In order to attempt to correct for and limit any attrition or nonresponse bias in the long-panel sample, an additional adjustment was performed. The adjustment was performed via a modelled response propensity using a logit model:

$$ln\left[\frac{p(x_i)}{1-p(x_i)}\right] = \alpha + \beta_1 \boldsymbol{X_i}$$

where $p(x_i)$ is the probability that household i was selected for the long-panel sample *and* successfully interviewed and $\boldsymbol{X}_i$ is a vast set of household characteristics from wave 1 which are potentially correlated with the probability of response[4]. It is necessary to use characteristics observed in wave 1 due to the panel nature of the sample. Adjustment using more recent household characteristics would contaminate the sample composition adjustments with dynamic changes that have occurred since the sample was first visited in wave 1. Following estimation of the logit model, the predicted probability of each household being interviewed, $\hat{p}(x_i)$, is derived using the model parameters. In principal, the inverse of the predicted probability could serve as the adjustment factor applied to the weights. However, in order to reduce potential distortions due to outliers, all units were sorted into 10 response classes (according to the deciles of $\hat{p}(x_i)$) and the average predicted probability within each class taken as

[4] The set of characteristics included in the response propensity model include

the adjustment factor. The adjustment factor is then applied to the existing weights. The adjusted weight was then winsorized at the 2nd and 98th percentiles to reduce the impact of outlier weights following the adjustments.

5. The weights were calibrated to reflect the distribution of the underlying population. The weights were calibrated to (1) reflect the total number of households in each Zone in 2010 (i.e. during the first wave of the GHS-Panel)

The cross-section and long-panel weights can be found in the cover page data files for both the post-planting (***secta_plantingw4.dta***) and post-harvest (***secta_harvestw4.dta***)**.** The variable names in both data files are ***wt_wave4*** for the cross-section weights and ***wt_longpanel*** for the long-panel weights.

4.0 Training of Field Staff and Data Entry Operators for the Survey

4.1 Training Design

Two levels of training were mounted for both the post-planting survey and the post-harvest survey. The first level was organized at NBS Headquarters in Abuja and was called the Training of Trainers (TOT). The participants in the TOT became the resource persons for the next level of training. The top management staff of the survey team participated in the TOT, which lasted for six days in the case of the post-planting and seven days for the post-harvest. The core training materials for the 2nd level training were harmonized and finalized during the TOT. The persons trained in the TOT were then sent to carry out the second level training.

The second level training for the post-planting visit was conducted over a ten-day period, while that of the post-harvest visit lasted for 13 days. In both visits' trainings, substantial amount of time were dedicated to theory including Survey Solutions CAPI training and two days to field practice and review. Participants in the training were Zonal Controllers, State Officers, Field Supervisors, Field Interviewers, and Data Editors (also called Survey Solutions Supervisors). Training instructions were given to the field staff by the resource persons from the management team (NBS) with support from the World Bank technical missions. About four or five resource persons were sent to each training centre to perform the training.

Specifically, the training consisted of (i) classroom instructions on the questionnaire, concepts and definitions, (ii) interview techniques, (iii) methods and field practices in performing actual interviews to ensure that field interviewers fully understood the questionnaire (iv) Survey Solutions CAPI, and (v) data review/checks and data management. In addition, participants did actual interviews in the field with households that were not scheduled to be part of the actual survey sample. Most of the training instructions are detailed in the interviewer's and supervisor's manuals which are also available.

4.2 Training Locations

Two geographic locations were used for the ToE – Keffi in Nassarawa state and Ibadan in Oyo state. The Keffi training centre housed the northern states (North Central, North East and North West), while the southern states (South East, South South and South West) had their training at the Ibadan training centre. Due to the large number of participants, each training centre had two halls where half of the participants were in one hall while the other half was in the other.

4.3 Evaluation of Field Personnel and Data Editor

At the end of the training session, trainees were assessed according to both a test that was administered on the material covered in the training process, and an evaluation by the resource

persons. Based on the results of the tests some interviewers and data editors were removed from the survey.

5.0 Field Work

5.1 Organization of Fieldwork

Data were collected by teams consisting of a supervisor, between 2 and 5 interviewers. The number of teams varied from state to state depending on the sample size or number of EAs selected. The teams moved in a roving manner and data collection lasted for between 20 – 30 days for each of the post-planting and post-harvest visits. Additional details on the structure of the visits are available in Section 7.

The GHS-Panel Wave 4 was administered in two visits: post-planting (July - September 2018) and post-harvest (January - February 2019). A tracking phase was conducted after both visits in October-November 2018 and April 2019 to interview long panel households that had moved from their location in the previous visit (Wave 3 post-harvest) or any household that had moved between the post-planting and post-harvest visits in Wave 4.

5.2 Gift to Households

As a show of appreciation for the panel households continued participation, all households that were located, were given a gift (even if they refused to participate). Gifts were given during both the post-planting and post-harvest visits and consisted of either a torchlight or a food flask (container for storing food). Households were very appreciative of the gifts and in many cases were essential to ensure continued participation of the household in the panel.

5.3 Pre-loaded Information

Basic information on every household was pre-loaded in the CAPI assignments for each interviewer. The information was pre-loaded to (1) assist interviewers in locating and identifying the household and (2) ensure that each pre-loaded unit (person, plot, etc.) is properly addressed and easily matched to previous visits. Basic household information (location, household head name, phone number, etc.) was pre-loaded in both visits for both refresh and long-panel households. For refresh households, the pre-loaded household information was taken from the listing exercise.

In the post-planting visit, the list of individuals from Wave 3 and their basic characteristics were uploaded for long panel households. This helped maintain the panel of individuals and ensure the status of each individual is confirmed in Wave 4. In the post-harvest visit, several information was pre-loaded that was collected in the post-planting visit. This includes (1) the list of individuals and their basic characteristics, (2) the list of plots of land and their basic characteristics and (3) the list of crops cultivated on each plot of land.

5.4 Fieldwork Monitoring and Evaluation

As an additional aid to ensuring good quality data, extensive monitoring was done of the field work. There were three levels of monitoring and evaluation. The first level of monitoring followed immediately after the zonal training. One (1) monitor was assigned to 2 states and all states were covered, including Federal Capital Authority, Abuja (FCT, Abuja). This monitoring was carried out by the technical team from the zonal training (i.e. the trainers) which included individuals from the Head Office of NBS. The second monitoring was carried out by NBS state officers and zonal controllers and took place over an extended period during the fieldwork. The third and final monitoring took place no later than a week before the end of fieldwork. The team involved in the third monitoring was selected from the team that carried out the first monitoring.

During first and second monitoring, the monitors made sure that proper compliance with the procedures as contained in the manual were followed, effected necessary corrections and tackled problems that arose. Where problems were found, these were corrected either directly or through a revisit to the household for verification of information or for further information.

6.0 Household Tracking Exercise

There were two separate tracking exercises conducted in Wave 4. The first was conducted directly following the post-planting visit and the second following the post-harvest visit. During the post-planting tracking exercise, households from the original sample that moved since the last time they were interviewed (either Wave 1, Wave 2 or Wave 3) and the first visit of Wave 4 were tracked. During the post-harvest tracking exercise, households (both original panel and newly refreshed sample) that moved between the post-planting and post-harvest visits of Wave 4 were tracked. During the main interview period of the post-harvest and post-planting visits, interviewers were instructed to complete a tracking form for all households who had relocated. In the case of households that moved to nearby locations, i.e. within the enumeration area, the interviewers were instructed to locate these households and administer the questionnaires.

6.1 Tracking States and Staff Assignments

Both tracking exercises were conducted by staff of the panel management team with support from interviewers in each of the applicable states. In states with two or less households to be tracked, the tracking was conducted by state staff only. Tables 6.1 and 6.2 below show the states where the tracking exercises took place, the number of households to be tracked and the number of field staff that were engaged in the activity.

Table 6.1: Number of Households to be Tracked and Allocation of Field Staff
POST PLANTING TRACKING

STATE WHERE HOUSEHOLD RELOCATED	NUMBER OF HOUSEHOLDS	VISITED BY HQ STAFF	NUMBER HQ PERSONS	NUMBER STATE PERSONS
ADAMAWA	1	NO		2
AKWA IBOM	2	NO		2
ANAMBRA	8	YES	1	1
BENUE	1	NO		2
CROSS RIVER	1	NO		2
DELTA	3	YES	1	1
EBONYI	3	YES	1	1
EKITI	4	YES		2
ENUGU	3	YES	1	1
IMO	1	NO		2
JIGAWA	1	NO		2
KATSINA	1	NO		2
KOGI	2	NO		2
KWARA	2	NO		2
LAGOS	9	YES	1	1

OGUN	3	YES	1	1
ONDO	1	NO		2
OSUN	2	NO		2
OYO	4	YES	1	1
RIVERS	2	NO		2
FCT	2	NO	1	1
TOTAL	**56**		**8**	**34**

Table 6.2: Number of Households to be Tracked and Allocation of Field Staff
POST-HARVEST TRACKING

STATE WHERE HOUSEHOLD RELOCATED	NUMBER OF HOUSEHOLDS	TO BE VISITED BY HQ STAFF	NUMBER OF HQ PERSONS	NUMBER OF STATE PERSONS
ADAMAWA	2	NO		2
AKWA IBOM	3	YES	1	1
ANAMBRA	7	YES	1	1
BAYELSA	2	NO		2
BENUE	2	NO		2
BORNO	1	NO		2
CROSS RIVER	1	NO		2
DELTA	3	YES	1	1
EDO	1	NO		2
EKITI	4	YES	1	1
ENUGU	4	YES	1	1
JIGAWA	2	NO		2
KADUNA	3	YES	1	1
KOGI	1	NO		2
LAGOS	6	YES	1	1
OGUN	5	YES	1	1
ONDO	6	YES	1	1
OYO	11	YES	1	2
RIVERS	9	YES	1	2
FCT	1	NO		2
TOTAL	**74**		**11**	**31**

6.2 Training of Tracking Staff

Training for both tracking exercises was conducted at the NBS head office for panel staff that would be involved in the tracking activity. The headquarters persons trained were to train their partner staff as well as state officers in their assigned state. A number of trainers also had the responsibility of training staff from states where no headquarters staff were slated to visit. The tracking fieldwork for the post-planting occurred from October 16 to November 4, 2019, while the post-harvest tracking was fielded over the period April 1 - 16, 2019.

6.3 Tracking Methodology

The tracking of households included the following steps:

- Discussion of the set of tracking households with relevant GHS-Panel interview team to obtain all information necessary. Use this information to finalise the list of households that will be tracked
- In order to properly prepare for the tracking field activities, the tracking exercise was initiated by the panel management team while at NBS head office. Contact was made with most of the households to be tracked by using the phone numbers given in the tracking questionnaire. Information was also used from the contact information on the questionnaires. That is, where households could not be contacted using the information in the tracking questionnaire, the contact information for family, friends and neighbours which was collected in the household questionnaire were also used. These preliminary tracking activities proved to be a very useful exercise in confirming the location of the relocated household and laying out the plan for the tracking fieldwork. Also, the opportunity was taken during the preliminary exercise, to obtain directions to households' new addresses and to set appointments for the interview.
- In cases where there was no useful phone information (either in the tracking questionnaire or household contact information) and the new address of the household was not known, the original location of the household was visited and effort made to obtain phone numbers for the household or the address. When information on the address was obtained, the household was visited by the team in charge of the state to which the household had moved.
- The required questionnaires were then administered to the household by the tracking team upon locating the household and securing their cooperation.

6.3.1 Tracking Households with Unknown Locations

Households with an unknown new address have been included as a part of the state in which they were originally located. It was the responsibility of the head office staff going to the original state of these "unknown" households to make an effort to gather further information on the place to which the household had relocated. This effort was made prior to the head office staff member's visit to the state. In cases where the household had moved to a new state, the new household location was passed to the staff member visiting the state to which the household had relocated. This household then became a part of that staff member's tracking assignment. If the

household had moved to another location within the original state, then it was included as a tracking assignment of the head office staff member in that state.

6.4 Identifying Tracked Households in the Data

Tracked households are identified by the *tracked_obs* variable that is included in the cover sheet data set. For households interviewed during the post-planting tracking exercise, see *tracked_obs* in the data file *secta_plantingw4* found in the Post-Planting Household data folder. For households interviewed during the post-harvest tracking exercise, see *tracked_obs* in the data file *secta_harvestw4* found in the Post-Harvest Household data folder.

7.0 Data Management and Description of Datasets

7.1 Data Management

7.1.1 CAPI

For the first time in GHS-Panel, the Wave four exercise was conducted using Computer Assisted Person Interview (CAPI) techniques. All the questionnaires, household, agriculture and community questionnaires were implemented in both the post-planting and post-harvest visits of Wave 4 using the CAPI software, Survey Solutions. The Survey Solutions software was developed and maintained by the Survey Unit within the Development Economics Data Group (DECDG) at the World Bank. Each enumerator was given tablets which they used to conduct the interviews. Overall, implementation of survey using Survey Solutions CAPI was highly successful, as it allowed for timely availability of the data from completed interviews.

7.1.2 Data Communication System

The data communication system used in Wave 4 was highly automated. Each field team was given a mobile modem allow for internet connectivity and daily synchronization of their tablet. This ensured that head office in Abuja has access to the data in real-time. Once the interview is completed and uploaded to the server, the data is first reviewed by the Data Editors. The data is also downloaded from the server, and Stata dofile was run on the downloaded data to check for additional errors that were not captured by the Survey Solutions application. An excel error file is generated following the running of the Stata dofile 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 is done on a daily basis throughout the duration of the survey, both in the post-planting and post-harvest.

7.1.3 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 (Headquarters in Survey Solutions). 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. The agriculture data were also checked to ensure that the plots identified in the main sections merged with the plot information identified

in the other sections. 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.

7.2 Description of Datasets

The GHS-Panel Wave 4 was administered in two visits: post-planting (July - November 2018) and post-harvest (January - April 2019). During each visit two questionnaires were administered to the household respondents (Household Questionnaire and Agricultural Questionnaire) and a third questionnaire was administered at the level of the enumeration area (Community Questionnaire). The tracking phases were completed in October/November 2018 (post-planting) and April 2019 (post-harvest). The tracking data is integrated into the post-planting and post-harvest structure, even though the data were actually collected in the tracking phase. The questionnaires implemented for tracking households were identical to those used in the main interview phase.

7.2.1 Household Data

In the Household Questionnaire, some of the modules were administered in both the post planting and post-harvest visit and others were only administered during one of the two visits. This should be taken into account when using the datasets.

Group 1: These modules are administered in both visits. For these topics we have complete information at two points in time during the year of the survey.

- Household Roster
- Labour
- Meals Away from Home
- Food Consumption and Expenditure
- Non-food Expenditure
- Food Security
- Contact Information

Group 2: These modules only appear in either the post-planting or the post-harvest visit

- Post-planting only
  - Savings and Insurance
  - ICT – Mobile Phone Banking
  - Credit
  - Household Assets
  - Housing
- Post-harvest only
  - Education
  - Health
  - Remittances
  - Nonfarm Enterprise and Income Generating Activities
  - Aggregate Food Consumption
  - Other Household Income
  - Safety Nets
  - Economic Shocks

Table 7-1 and Table 7-2 show the sections of the Household Questionnaire and their corresponding datasets.

Table 7-1: Post-planting Household datasets

Section	Section Name	Dataset Filename
Cover	Cover	secta_plantingw4
1	Household Roster	sect1_plantingw4
3A	Labour	sect3_plantingw4
		sect3q38_plantingw4
3B	Time Use	sect3b_plantingw4
4A	Savings and Insurance	sect4a1_plantingw4
		sect4a2_plantingw4
4B	ICT – Mobile Phone Banking	sect4b_plantingw4
4C	Credit	sect4c1_plantingw4
		sect4c2_plantingw4
		sect4c3_plantingw4
5	Household Assets	sect5_plantingw4
7A	Meals Away from Home	sect7a_plantingw4
7B	Food Expenditure	sect7b_plantingw4
8	Non-Food Expenditures	sect8a_plantingw4
		sect8b_plantingw4
		sect8c_plantingw4
9	Food Security	sect9_plantingw4
11	Housing	sect11_plantingw4

Table 7-2: Post-harvest Household datasets

Section	Section Name	Dataset Filename
Cover	Cover	secta_harvestw4
1	Household Roster	sect1_harvestw4
2	Education	sect2_harvestw4
3	Labour	sect3a_harvestw4
		sect3b_harvestw4
4	Health	sect4a_harvestw4
6	Remittances	sect6_harvestw4
9	Non-farm Enterprises and Income Generating Activities	sect9a_harvestw4
		sect9b_harvestw4
		sect9c_harvestw4
		sect9d_harvestw4
10A	Meals Away From Home	sect10a_harvestw4
10B	Food Expenditures	sect10b_harvestw4
10C	Aggregate Food Consumption	sect10c_harvestw4
11	Non-food Expenditures	sect11a_harvestw4
		sect11b_harvestw4
		sect11c_harvestw4
		sect11d_harvestw4
12	Food Security	sect12_harvestw4
13	Other Household Income	sect13_harvestw4
14	Safety Nets	sect14a_harvestw4
		sect14b_harvestw4
15A	Economic Shocks	sect15a_harvestw4

7.2.2 Agriculture Data

It should be noted that in the Agriculture Questionnaire, the plot roster and land inventory information collected during the post-planting visit is updated during the post-harvest visit in the Land section to identify plots they have disposed of since the first, post-planting visit.[5] As with the Household Questionnaire, some modules were administered in both visits. For these modules, during the post-harvest visit, information was gathered on the activities since the post-planting interview.

Table 7-3 and Table 7-4 show the sections of the Agriculture Questionnaire and their corresponding datasets.

[5] In theory, some plots in Wave 4 can be matched to Wave 1, 2 and 3 using the characteristics of the plots. However, the plot description and codes were not prefilled from previous waves. Thus plots cannot be matched across plots using plot IDs.

Table 7-3: Post-planting Agriculture datasets

Section	Section Name	Dataset Filename
11A	Plot Roster	sect11a_plantingw4
		sect11a1_plantingw4
11B1	Land Inventory	sect11b1_plantingw4
		sect11b1q10_plantingw4
11C1	Household Labour Hired and Exchange Labour	sect11c1a_plantingw4
		sect11c1b_plantingw4
11E	Seed acquisition	sect11e1_plantingw4
		sect11e2_plantingw4
11F	Planted Crops	sect11f_plantingw4
11I	Animal Holdings	sect11i_plantingw4
11J	Animal Costs	sect11j_plantingw4
11K1	Animal Power and Dung	sect11k1_plantingw4
11K2	Milk Production	sect11k2_plantingw4
11K3	Egg Production	sect11k3_plantingw4
11L1	Extension Services	sect11l1_plantingw4
11L2	Extension Sources	sect11l2_plantingw4

Table 7-4: Post-harvest Agriculture datasets

Section	**Section Name**	**Dataset Filename**
A1	Land	secta1_harvestw4
A2a	Household Labour	secta2a_harvestw4
A2b	Hired and Exchange Labour	secta2b_harvestw4
11C2	Input Use on Plot	secta11c2_harvestw4
11C3	Input Acquisition	secta11c3_harvestw4
		secta11c3q12_harvestw4
A3i	Field Crop Harvest	secta3i_harvestw4
A3ii	Crop Disposition	secta3ii_harvestw4
A3iii	Tree and Permanent Crop Harvest	secta3iii_harvestw4
PHL1	Post-Harvest Labour - Household	sectaphl1_harvestw4
PHL2	Post-Harvest Labour – Hired and Exchange	sectaphl2_harvestw4
A4	Agricultural Capital	secta4_harvestw4
A5b	Extension Services	secta5b_harvestw4
A9a	Fishing	secta9a_harvestw4

7.2.3 Community Data

Table 7-5 and Table 7-6 show the sections of the Community Questionnaire and their corresponding data sets.

Table 7-5: Post-planting Community datasets

Section	Section Name	Dataset Filename
Cover	Cover	sectc_plantingw4
C1	Respondents Characteristics	sectc1_plantingw4
C2	Food Prices	sectc2_plantingw4
C3	Labour	sectc3a _plantingw4
		sectc3b_plantingw4
C4	Land Prices and Credit	sectc4a_plantingw4
		sectc4b_plantingw4
		sectc4c_plantingw4

Table 7-6: Post-harvest Community datasets

Section	Section Name	Dataset Filename
Cover	Cover	sectc_harvestw4
C1	Respondents Characteristics	sectc1_harvestw4
C2	Community Infrastructure and Transportation	sectc2_harvestw4
C3	Community Organizations	sectc3_harvestw4
C4	Community Resource Management	sectc4_harvestw4
C5	Community Changes	sectc5_harvestw4
C6	Community Key Events	sectc6_harvestw4
C7	Community Needs, Actions, and Achievements	sectc7_harvestw4
C8	Food Prices	sectc8_harvestw4
C9	Labour	sectc9a_harvestw4
		sectc9b_harvestw4

7.2.4 Confidential information

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

7.2.5 Auxiliary information

Two additional data sets are released with Wave 4 which includes auxiliary information of the interviews: ***aux_plantingw4.dta*** and ***aux_harvestw4.dta***. These include the starting and the ending time of the interviews, respondents of the individual section of the questionnaire.

7.2.6 Non-Standard Units Conversion Factors

Food and crop 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), two types of conversion factor are included in the datasets. The first is the conversion factors for food quantities in the food consumption file. The second is the conversion factors for crops to be used with the agricultural module. For more information on these files and how to use them, see Section 8.3.1.

7.2.7 Geospatial Variables

To increase the use of the GHS-Panel data, a set of geospatial variables has been provided by using the georeferenced plot and household locations in conjunction with various geospatial databases that were available to the survey team. More information is available in Appendix 3 on how these variables are constructed and linked to the GHS-Panel data. The table in Appendix 3 provides the name, type, source, reference period, resolution, description, and source of each geospatial variable included.

7.2.8 Consumption Expenditure Data

One of the objectives of the GHS-Panel is allow for detailed household welfare analysis across multiple dimensions of wellbeing. In order to arrive at a consistent measure of well-being, the GHS-Panel collects all the necessary information to compute the prevailing measure of well-being in Africa: household consumption expenditures. Aggregate consumption expenditures were calculated from the GHS-Panel data. The final adjusted measure of aggregate consumption expenditure (***totcons_adj_norm***) as well as the components of expenditure are contained in the data file ***totcons_final.dta*** provided along with the GHS-Panel microdata files. The calculation of aggregate consumption expenditures is a complicated process with many steps, assumptions and adjustments to arrive at the final estimate. Appendix 4 provides full details on how aggregate consumption expenditures were computed from the GHS-Panel data.

8.0 Using the Data

8.1 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 5a and question 5b, a third variable, question 5c, 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.

Every effort was made to keep question numbers (and thus variable names) as consistent as possible with wave 3. If questions were dropped in Wave 4, the numbering was preserved. If questions were added in the middle of a section, a letter was added to the question number at that space in the sequence (e.g. if added before question 2, the question number would be 2a). This was done to make utilization of the data sets across the three waves as consistent as possible.

8.2 Merging Datasets

8.2.1 Household and Agriculture Datasets

All household and agriculture datasets in both the post-planting and post-harvest files 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 type files, both the variables ***hhid*** and ***indiv*** would be used. In the agriculture datasets, plot files are merged using ***hhid*** and ***plotid*** while crop files are merged using ***hhid, plotid*** and ***cropid*****.**

8.2.2 Post-Planting and Post-Harvest Datasets

Post-planting and post-harvest files can be merged using the methodology explained above. That is, the ***hhid*** is the same for a specific household in the post-planting and post-harvest visit. It should be noted that there was some attrition of households between the post-planting and post-harvest visits so some households in the post-planting files will not have a match in the post-harvest data sets. Note also that people may have left the households or joined them in the time between the two visits. Thus the number of people per household will vary between visits.

8.2.3 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. Merging of community files within the round or with community files from the other round or with any of the household or agriculture files from either round should be done using the ***lga*** and ***ea*** variables, in that order.

8.3 Food and Crop Unit Measures

When collecting information on food or crop quantities (e.g. amount of food consumed, amount of crop harvested, 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 Wave 4, the unit list was expanded 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.

8.3.1 Unit Conversion Factors

In order to utilise and compare quantities in different units they must be converted into a common unit using conversion factors. Prior to Wave 3, a market survey was conducted 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 difference in units and conversion across the country. Where there were sufficient observations, Zone-level conversions were calculated. In Wave 4, a small number of additional conversions were also collected for additional units not previously included.

In Wave 3, the conversion factors were provided in separate data files that accompanied the main data. This required the data user to merge these conversions into the relevant data files where conversion needed to be applied. In Wave 4, these conversions are already provided in the relevant data files. For example, the dataset ***sect7b_plantingw4.dta*** contains question 2, which asks how much the household consumed of each food item. Question 2 has four main components: the quantity, the unit, the size (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 ***s7bq2a*s7bq2_cvn***.

Crop harvest conversion factors have a slightly different treatment from the other conversion factors in the GHS-Panel data. Conversion for some crops have been further adjusted based on the "condition" of the crop being reported. The condition in this case refers to the different forms that a crop can take. For example, maize can come unshelled (i.e. maize cobs) or shelled (i.e. maize grains removed from the cob). Respondents were asked to further specify the condition of the quantity of the crop they are reporting. In order to make quantities comparable, the conversions have been adjusted to account for the different conditions. In all cases, the conversions are adjusted to give the shelled equivalent quantity for unshelled crop. This adjustment was also applied for standard units like kilograms. For example, the conversion for one kilogram of *unshelled* maize will not be 1:1 since 1 kilogram of unshelled maize is not equal to 1 kilogram of shelled maize.

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.

8.3.2 Reference Photo Album

In Wave 3 the collection of photos was greatly expanded and improved. These reference photos were also used during Wave 4 data collection. The photos were collected in a systematic manner during the market survey prior to Wave 3 where the item-unit weights were also collected. During the market survey, interviewers were instructed to follow strict protocols when taking the photographs 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 noncontainer 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 forthcoming guidebook produced by the 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*.

9.0 Overall Problems and Challenges Faced During Wave 4

Designing and implementing a complex survey such as the GHS-Panel presents various challenges. In this section we outline some key issues that arose, lessons learned and make recommendations for the next wave of the survey.

9.1 GPS Measurement of Plots

During the post-planting visit, there were some challenges with measurement of plots using GPS devices. Only about 80 percent of plots were measured. Non-measurement was concentrated in several States. In some cases, this was the result of flooding (especially in South East), but in other cases the interviewers were reluctant to measure plots that were distant from the household (though still within the Local Government Area). During the post-harvest visit, interviewers were asked to measure plots that were not measured in the post-planting visit. Additional questions were added in the post-harvest agricultural questionnaire (*Section A1: Plot Roster*) and the measurement status of each plot was included in the plot roster prefilling form.

The majority of these plots were measured in the post-harvest visit. This increased the number of GPS measured plots to about 90 percent. Some plots could still not be measured due to flooding or security concerns in some areas.

9.2 Security Problems

The most significant challenge faced during Wave 4 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 road blocks 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. Sometimes, the teams had to adjust their plans based on the situation on the ground.

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. The remote data quality monitoring framework will be further strengthened in future waves, though hopefully the security situation will have improved to allow direct monitoring.

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 LSMS-ISA website:

http://surveys.worldbank.org/lsms/programs/integrated-surveys-agriculture-ISA

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 presentation to the National Bureau of Statistics (see address below) and the Data Production and Methods Team of the World Bank (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

LSMS Database Manager
Data Production and Methods Team (DECPM)
The World Bank
1818 H Street, NW
MSN MC3-306
Washington, DC 20433
http://surveys.worldbank.org/lsms
Email: lsms@worldbank.org

Appendix 2: Agriculture Land Conversion Factors

The table below shows the conversion factors used to convert self-reported land areas (for agricultural land area of crops planted and harvested) into hectares.

General Conversion Factors to Hectares

Zone	Unit	Conversion Factor
All	Plots	0.0667
All	Acres	0.4
All	Hectares	1
All	Sq Meters	0.0001

Zone Specific Conversion Factors to Hectares

Zone	Conversion Factor		
	Heaps	Ridges	Stands
North Central (1)	0.00012	0.0027	0.00006
North East (2)	0.00016	0.004	0.00016
North West (3)	0.00011	0.00494	0.00004
South East (4)	0.00019	0.0023	0.00004
South South (5)	0.00021	0.0023	0.00013
South East (6)	0.00012	0.00001	0.00041

Note: All conversion is to Hectares

Appendix 3: Geospatial Variables[6]

To enhance the GHS-Panel data, a set of geospatial variables has been generated using the georeferenced plot and household locations in conjunction with various geospatial databases that were available to the survey team. These include simple measures of distance, climatology, soil and terrain and other environmental factors. The variables are intended to provide some understanding of how geophysical characteristics vary across households and between communities. All geospatial variables in the public release have been produced using anonymized location data (see below for description of anonymization method).

NGA_HouseholdGeovariables_Y4

The household-level file, *NGA_HouseholdGeovariables_Y4,* contains a range of variables measuring (on the basis of the anonymized EA location, or anonymized household location for movers) distance to other features, climatology, landscape typology, soil and terrain, and growing season parameters. The observations are uniquely identified by **hhid**.

This file also contains the modified GPS coordinates, which enables users to generate their own spatial variables while preserving the confidentiality of sample household and communities. Following the method developed for the Measure DHS program, the coordinate modification strategy relies on random offset of cluster center-point coordinates (or average of household GPS locations by EA in GHS-Panel) within a specified range determined by an urban/rural classification. For urban areas a range of 0-2 km is used. In rural areas, where communities are more dispersed and risk of disclosure may be higher, a range of 0-5 km offset is used. An additional 0-10 km offset for 1% of rural clusters effectively increases the known range for all rural points to 10 km while introducing only a small amount of noise. Offset points are constrained at the state level, so that they still fall within the correct state for spatial joins, although boundary precision may be an issue for clusters located very close to the border.

In the third wave of panel data collection some households are tracked to a new location. These include both local and long-distance moves, although a majority of tracked households are within 5 km of the original location. The public coordinates for new locations that are within 5 km of the original household location remain unchanged (modified coordinates of original sample EA). The public coordinates of tracked households that are more than 5 km from original location are assigned a new offset location, according to the method described above. Additionally, the distance from original location is provided for tracked households with new locations.

The result is a set of coordinates, representative at the cluster level, that fall within known limits of accuracy. Users should take into account the offset range when considering different types

[6] Users have occasionally requested actual geographic locations of households in the sample from the LSMS Office. The World Bank is not authorized to release these data. All requests for actual geographic locations must be made to NBS.

of spatial analysis. Analysis of the spatial relationships between locations in close proximity would not be reliable. However, spatial queries using medium or low-resolution datasets should be minimally affected by the offsets. Zonal statistics (average or range of values within an area corresponding to the known range) could help minimize the effect of offsets when combining with large scale data or high-resolution grids with a high degree of local variation.

The tables below provide the name, type, source, reference period, resolution, and description of each variable. With the exception of 3 distance variables (dist_road2, dist_popcenter2, dist_borderpost2), the source spatial data are same as used in baseline survey year. The three distance variables have been updated using more reliable spatial datasets. In addition, the phenological variables have been replaced with data from the NOAA Climate Data Record NDVI dataset, and the rainfall variables are based on the Africa Rainfall Climatology (ARC2) dataset, which has a longer length of record but is largely consistent with RFE.

NGA_PlotGeovariables_Y4

The household plot-level file, *NGA_PlotGeovariables_Y4,* contains four variables measuring plot distance to household, slope of plot, elevation of plot and plot potential wetness index. The observations are uniquely identified by the combination of **hhid plotid**. The observations included in this file are plots that are owned and/or cultivated by the household and that have been visited for GPS-based land area measurement.

Coordinates of the plots are not included.

Table A4.7 NGA_HouseholdGeovariables_Y4

Theme	Source	Dataset Title	Variable Name	Variable Type	Reference Period	Resolution	Description	Web
	FERMA	Household Distance to Main Road	dist_road2	Continuous	2013	N/A	Distance to nearest federal road included in FERMA survey, 2013	
	WorldCities	Household Distance to Towns	dist_popcenter2	Continuous	2012	N/A	Population for cities of > 20,000 listed in worldcities database, c. 2012	http://www.worldcities.us/nigeria_cities/
	USAID FEWSNET	Household Distance to Key Market Centers	dist_market	Continuous	N/A	N/A	Household distance to nearest major market (FEWSNET key market centers)	
	GoogleEarth and other map sources	Household Distance to Border Posts	dist_borderpost2	Continuous	N/A	N/A	Household distance to nearest border post on main road, primary crossings only	
	Wikipedia and other map sources	Household Distance to State Capital	dist_admctr	Continuous	N/A	N/A	Household distance to to the capital of the State of residence	
Climatology	UC Berkeley	WorldClim Bioclimatic Variables	af_bio_1	Continuous	1960-1990	0.008333 dd	Average annual temperature calculated from monthly climatology, multiplied by 10 (°C)	http://www.worldclim.org/bioclim
	UC Berkeley	WorldClim Bioclimatic Variables	af_bio_8	Continuous	1960-1990	0.008333 dd	Average temperature of the wettest quarter, from monthly climatology, multiplied by 10. (°C)	http://www.worldclim.org/bioclim
	UC Berkeley	WorldClim Bioclimatic Variables	af_bio_12	Continuous	1960-1990	0.008333 dd	Total annual precipitation, from monthly climatology (mm)	http://www.worldclim.org/bioclim
	UC Berkeley	WorldClim Bioclimatic Variables	af_bio_13	Continuous	1960-1990	0.008333 dd	Precipitation of wettest month, from monthly climatology (mm)	http://www.worldclim.org/bioclim
	UC Berkeley	WorldClim Bioclimatic Variables	af_bio_16	Continuous	1960-1990	0.008333 dd	Precipitation of wettest quarter, from monthly climatology (mm)	http://www.worldclim.org/bioclim
Landscape Typology	IIASA-IFPRI	Cropland Map v8	hybrid_V8	Continuous	2009	0.008333 dd	Percentage cropland ranging from 0 to 100	http://ionia1.esrin.esa.int/
	WorldPop	WorldPop	popdensity	Categorical	2015	0.00833 dd	2015 Population Density (people per km2), from Global High resolution Population Denominators Project.	https://www.worldpop.org/geodata/summary?id=5164
	IFPRI	IFPRI standardized AEZ based on	ssa_aez09	Categorical		0.008333 dd	Agro-ecological zones created using WorldClim climate data and 0.0833dd resolution LGP data from IIASA.	http://harvestchoice.org/production/biophysical/agroecology

Theme	Source	Dataset Title	Variable Name	Variable Type	Reference Period	Resolution	Description	Web
		elevation, climatology						
Soil & Terrain	NASA	SRTM	srtm1k	Continuous		0.00833 dd	Elevation (m), aggregated to 1km block	ftp://xftp.jrc.it/pub/srtmV4/arcasci/
	USGS	Slope (percent)	slopepct	Continuous		0.008333 dd	Derived from 90m SRTM, aggregated to 1km block	http://pubs.usgs.gov/of/2007/1188/, data provided USGS upon request
	AfSIS	Topographic Wetness Index	twi	Continuous		0.000833 dd	Downloaded from AfSIS website. Derived from modified 90m SRTM. Local upslope contributing area and slope are combined to determine the potential wetness index: WI = ln (A s / tan(b)) where A s is flow accumulation or effective drainage area and b is slope gradient.	http://www.ciesin.columbia.edu/afsis/bafsis_fullmap.htm#
	FAO	Harmonized World Soil Database	SQ1	Categorical		0.083333 dd	Nutrient availability	http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/
	FAO	Harmonized World Soil Database	SQ2	Categorical		0.083333 dd	Nutrient retention capacity	http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/
	FAO	Harmonized World Soil Database	SQ3	Categorical		0.083333 dd	Rooting conditions	http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/
	FAO	Harmonized World Soil Database	SQ4	Categorical		0.083333 dd	Oxygen availability to roots	http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/

Theme	Source	Dataset Title	Variable Name	Variable Type	Reference Period	Resolution	Description	Web
	FAO	Harmonized World Soil Database	SQ5	Categorical		0.083333 dd	Excess salts	http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/
	FAO	Harmonized World Soil Database	SQ6	Categorical		0.083333 dd	Toxicity	http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/
	FAO	Harmonized World Soil Database	SQ7	Categorical		0.083333 dd	Workability (constraining field management)	http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/
Crop Season Parameters	NOAA CPC	ARC 2	anntot_avg	Continuous	1983-2012	0.1 dd	Long-term average annual total rainfall (mm)	ftp://ftp.cpc.ncep.noaa.gov/fews/fewsdata/africa/arc2
	NOAA CPC	ARC 2	wetQ_avg	Continuous	1983-2012	0.1 dd	Long-term average total rainfall in wettest quarter (mm) within 12-month period	
	NOAA CPC	ARC 2	wetQ_avgstart	Continuous	1983-2012	0.1 dd	Average start of wettest quarter in dekads 1-36, where first dekad of Jan =1	
	NOAA CPC	ARC 2	h2018_tot	Continuous	2018	0.1 dd	12-month total rainfall (mm), starting January 2018	
	NOAA CPC	ARC 2	h2018_wetQ	Continuous	2018	0.1 dd	Total rainfall in wettest quarter (mm) within 12-month periods starting January 2018	
	NOAA CPC	ARC 2	h2018_wetQstart	Continuous	2018	0.1 dd	Start of wettest quarter in dekads 1-36, where first dekad of January 2018 =1	
	NOAA CDR	NDVI version 4	ndvi_avg	Continuous	1983-2012	0.05 dd	Long-term average NDVI value in primary growing season (highest quarter)	https://www.ncei.noaa.gov/data/avhrr-land-normalized-difference-vegetation-index/access
	NOAA CDR	NDVI version 4	ndvi_max	Continuous	1983-2012	0.05 dd	Long-term maximum dekadal NDVI value in primary growing season (highest quarter)	

Theme	Source	Dataset Title	Variable Name	Variable Type	Reference Period	Resolution	Description	Web
	NOAA CDR	NDVI version 4	h2018_avg	Continuous	2018	0.05 dd	Average NDVI value in primary growing season (highest quarter) in 2018	
	NOAA CDR	NDVI version 4	h2018_max	Continuous	2018	0.05 dd	Maximum dekadal NDVI value in primary growing season (highest quarter) in 2018	

Appendix 4: Construction of the Consumption Aggregate

I. Objectives and the use of the consumption aggregate

Consumption aggregates serve as a foundation for measuring the welfare of the population. The Nigeria General Household Survey-Panel, 2018-19, Panel Wave 4 (GHS-Panel) collected all necessary data to calculate the required components of consumption aggregate. Construction of the welfare aggregate is a complex process, which combines the well-established procedures as well as value judgments, reflecting the country context. A complete discussion of theory and issues involved in construction of the consumption aggregate can be found in Deaton and Zaidi (Deaton & Zaidi, 2002), while this note summarizes the key steps and decisions taken while computing the aggregate using GHS 2018-19 data.

The second and more important objective of this note is to inform the users about proper use of the consumption aggregate based on GHS-Panel 2018-19. The data collection for GHS-Panel 2018-19 coincided with implementation of the Nigeria Living Standards Survey, 2018-19 (NLSS). While both surveys utilize similar survey instruments and most variables overlap, NLSS and GHS-Panel differ in terms of objectives, sampling and the structure of the survey instrument. In contrast to GHS-Panel, NLSS is a larger cross-sectional survey of more than 22,200 households specifically designed to measure poverty and welfare in one point of time. The GHS-Panel is a smaller panel survey of around 5,000 households, with a strong focus on agricultural activities. Data collection for the GHS-Panel is structured in two visits, corresponding to two agricultural seasons. NLSS is representative at state level, while the GHS-Panel is statistically valid at the larger, aggregated geo-zone level. These differences are also reflected in the values of the consumption aggregate, i.e. aggregates are not technically comparable and are not expected to converge.

It is important to note that while NLSS 2018-19 and GHS-Panel 2018-19 coincided in time, the data and consumption aggregate from the NLSS 2018-19 are the only consistent source of information to measure poverty in the country. Given the different design and objectives of the GHS-Panel, applying the consumption aggregate based on GHS-Panel 2018-19 to measure poverty in Nigeria would result in inconsistent estimates and is advised against. Thus, the GHS-Panel based consumption aggregate is to be only used for the welfare ranking of the households in relative terms, and not for measuring the poverty indicators in absolute terms. For the official measure of poverty in the country, the NLSS 2018-19 should be used[7].

II. Construction of the welfare aggregate

[7] For information on the NLSS 2018-19 as well as to request access to the microdata, please consult either the dedicated webpage on NBS's National Data Archive or on the World Bank's Microdata Library.

It is difficult to overstate the significance of properly constructed welfare aggregate, since consumption aggregate serves as a basis for estimating the poverty line(s) and the welfare indicators for the country.

The consumption aggregate is the monetary value of food and non-food goods and services consumed by the household. Corresponding to the available modules in GHS-Panel 2018-19 and practices adopted in the previous living standards surveys in Nigeria, the consumption aggregate consists of 6 primary components:

1. Food consumption, including purchased, self-produced and received as a gift food items;
2. Meals or expenditures on food outside of home;
3. Non-food section, which includes frequently purchased items; non-food goods and services purchased on a regular basis and non-food expenditures incurred on less regular intervals, e.g. on semi-annual or annual basis;
4. Education expenditures;
5. Health expenditures;
6. Dwelling or rent expenditures.

i. Food consumption expenditures and spending on meals

The nominal food consumption expenditures are calculated as the total value of consumed food products, obtained from various sources including from purchases, self-produced and received as gifts etc. The food consumption module of the GHS-Panel questionnaire asks if members of a household consumed a particular food item in the last 7-days and if so, the physical quantity consumed at the household-level. Households do not report the monetary value of food consumed, therefore, to obtain/calculate the value of consumption, one needs to utilize the purchase portion of the consumption module to impute the unit values of each food item.

The food consumption module covers 16 general food categories (grains, starchy roots and tubers, cereals, baked food, etc.), and more than 160 individual food items, including drinks and alcohol. The module includes questions on the sources of the consumed food: from purchases, self- produced and from other sources, like gifts, barter, in-kind payments, etc.

The estimation of the value of the food consumption is undertaken in three steps: 1) convert the consumed and purchased food quantities from non-standard units of measurement into standard units (e.g. from *mudu* to *kilograms*); 2) calculate the unit values, as a ratio of expenditures or cost of purchased food item to the standard quantity in kilograms; 3) then the unit values are used to obtain the value of food consumption for each item.

Households in Nigeria and across regions within the country use multiplicity of non-standard measurement units (e.g. pieces, bunches, heaps, bags, paint rubber, cups etc.) for different food items, which might affect the accuracy of food consumption. Within the context of NLSS and

GHS-Panel surveys the NBS team undertook a challenging task of creating high-quality conversion factors for each non-standard unit of measurement and for each food item, disaggregated by the regions where possible. For a very few non-standard units of measurement and items which lacked the conversion rate, the decision was taken to apply the conversion of similar food items, e.g. if conversion for heap of nuts was missing then conversion factor for heap of rice was used. Except for few cases, no treatment of outliers in quantities was applied. In the second step the unit values for each food item and each household were calculated as a ratio of cost of purchases of food item to the quantity purchased in kilograms in each month.

$$uv_{i,m}^{h} = \frac{expenditures_{i,m}^{h}}{quantity_{i,m}^{h}}$$

Where: *uv* stands for unit value; superscript *h* here and elsewhere in this note denotes the household; subscript *i* denotes specific item (here food item); *m* – specific month.

Given the importance of unit values in computing the food consumption, the efforts were made to identify and replace the outliers in unit values, i.e. values in log terms higher or lower than 2.5 standard deviation from the mean have been identified and removed.

$$outlier\left(uv_{i,m}^{h}\right) = abs\left[\frac{\ln\left(uv_{i,m}^{h}\right) - mean\left(ln\left(uv_{i,m}^{strata}\right)\right)}{std\left(ln\left(uv_{i,m}^{strata}\right)\right)}\right] > 2.5$$

If a household reported a purchase of the food item, then the unit value is available, and the unit values are taken at the household level. In cases when the unit value is an outlier or missing (i.e. self-produced or was received as a gift, etc.), i.e. not available, the unit values represent the median values for specific month at the lowest available strata level (i.e. hierarchical imputation procedure): first taken from the level of enumeration area, then at state, zone, and finally at country levels, in a given sequence:

$$uv_{i,m}^{h} = uv_{i,m}^{h}, \quad if\ uv_{i,m}^{h}\ is\ not\ missing\ or\ not\ outlier$$
$$uv_{i,m}^{h} = median\ (uv_{i,m}^{ea}), \quad if\ uv_{i,m}^{h}\ is\ missing\ or\ outlier$$
$$uv_{i,m}^{h} = median\left(uv_{i,m}^{state}\right), \quad if\ uv_{i,m}^{h}\ is\ missing\ or\ outlier$$
$$uv_{i,m}^{h} = median\left(uv_{i,m}^{zone}\right), \quad if\ uv_{i,m}^{h}\ is\ missing\ or\ outlier$$
$$uv_{i,m}^{h} = median\left(uv_{i,m}^{country}\right), \quad if\ uv_{i,m}^{h}\ is\ missing\ or\ outlier$$

In the third and final step, the unit values are multiplied by the quantity consumed (including from own-production, food received as a gift or for free, etc.). The food consumption aggregate for a household is then a sum of imputed values of food consumption, across all food items the household consumed in the last 7 days, which is then annualized.

$$x_{Food,m}^{h} = \sum_{i=1}^{I} quantity_{i,m}^{h,cons} * uv_{i,m}^{h} * \left(\frac{365}{7}\right)$$

The procedure for calculating the expenditures on meals consumed outside of home is different and more straightforward. The households directly report the amount spent on food consumed outside of home e.g. in cafeteria, streets, café and restaurants, by categories, like breakfast, lunch and dinner or snacks or drinks. Therefore, the monetary value of meals outside of home is a sum of all expenditures on meals for each household.

$$x^{h}_{Meals,m} = \sum_{i=1}^{I} spending\ on\ meal\ category^{h}_{i,m} * (\frac{365}{7})$$

ii. Non-food expenditures

The non- food component of the consumption aggregate includes a heterogenous and wide-ranging group of expenditures on non-food items:

(1) Recreational and culture items: newspapers, magazines, cinemas, tobacco, etc.

(2) Energy: petroleum, kerosene, liquid gas, diesel, charcoal, electricity, etc.

(3) Clothing and footwear

(4) Personal services: repairs of appliances, etc.

(5) Household goods and appliances: furnishing, small electric appliances, etc.

The non-food items that were included in the consumption aggregate are: i) not "lumpy", e.g. purchase of a car or long-term durable goods; ii) not used as inputs in the production or as investment; iii) not acquired for the purpose of transfer to other households.

Given that expenses for non-food items that were included generally take place with different frequencies, the questionnaire asked the household to recall their expenditure on these items using different recall periods: 7 day, the last month and the last year. For the non-food part of the consumption aggregate, the challenge is to decide what needs to be excluded. To avoid bias in ranking of households, non-food expenditures exclude infrequent expenses like mortgages and ceremonial spending (funerals, weddings, dowries etc.).

Non-food expenditures are valued at the purchase or self-reported acquisition value. Depending on the reference period the expenditures are annualized by factor of (365/7) for 7-day recall items or (12/1) for one month recall items, except for expenditures on dwelling repairs, which were included without annualization.

$$x^{h}_{Non-food,m} = \sum_{i=1}^{I} purchase\ of\ nonfood\ item^{h}_{i,m} * (appropriate\ annualiztion\ factor)$$

iii. Housing expenditures/ rent

Housing cost is defined as implicit value or benefit that household receives from occupying a dwelling and not the expenditures on purchasing the dwelling itself. To measure monetary flow of housing services, the concept of rent is the appropriate choice. Since most households own the dwelling and thus do not explicitly pay the rent, a hedonic regression model is applied to estimate coefficients and predict the rent. The dependent variable is actual rent paid in log form, regressed on a set of housing variables/attributes like, location, number of rooms, material of roof, material of floor, material of wall, amenities/utilities (toilet, water sources, garbage collection, etc.):

$$ln(R^h) = a_0 + \sum_{n=1}^{N} f_n(X_n^h)\, b_n^h + e^h$$

Where:

R^h is the actual rent;

X_n^h is observable housing attributes/characteristics (n).

The imputed rent is a predicted value of housing from regression, that was re-transformed into Naira terms from log form, using Duan-Smearing transformation method (Duan, 1983) and applied to the households that own their dwelling or do not pay rent for the dwelling they reside in. The challenge in estimating the rent is that rental market in some areas specifically in rural areas is quite thin. To overcome this, the hedonic model is applied at more aggregated strata levels, instead of state levels. Also given some housing characteristics, the model results in extremely low rent values in rural areas. To avoid this the minimum value of rent is limited from below at the minimum of actual rent in a specific location/stratum.

iv. Education expenditures

Education expenditures is important component of the consumption aggregate. While spending on education could be viewed as long-term investment and thus be excluded from the aggregate, it is also true that education directly contribute to the household welfare. Since, education expenditures are regular across all groups of households and there is positive relationship between school spending and consumption deciles, education expenditures are included in the aggregate.

Expenditures on education includes all school related expenses from pre-school to tertiary education: school fees, uniform, textbooks, meals and lodging, transport, gifts to teachers and services to school, private tutoring and other expenses for education and schooling. Education expenses were recorded for past school/academic year and no annualization factor was applied.

$$x^{h}_{Education,m} = \sum_{i=1}^{I} actual\ school\ related\ spending\ category^{h}_{i,m}$$

v. Health care expenditures

The health module of GHS-Panel extensively covers questions, reflecting health care seeking behaviour of households. Health expenditures are recorded on consultations, medicines, laboratory exams, hospitalization charges, transport and other out of pocket costs related to health care. Elsewhere the motivation for excluding the health-related expenditures from the welfare aggregate is related to consideration of health cost as a "regrettable necessity". If a member of household falls ill and incurs medical expense this will increase total expenditures and therefore household's level of welfare when in fact, the opposite may be the case. In similar vein, some components of health expenditure have the characteristic to be a reaction to a shock, for which extraordinary means are used.

However, the analyses of the data showed that health expenditures should be included in the consumption aggregate. First the prevalence of health insurance in Nigeria is negligible, below 0.5 percent of population. Thus, access to health care is welfare enhancing and affects ranking of households. Second, close to 80 percent of health cost is attributed to the purchase of drugs and medicine, which are not directly subsidized. Again, excluding these costs will negatively affect the welfare ranking. Finally, the health expenditures are prevalent across all income groups and the share of health cost in total household budget is significant. There is a positive correlation between healthcare spending and consumption deciles, pointing to significant elasticity of health expenditures with respect to total expenditures. This clearly implies that health is central component of household welfare, as well as important area of public policy.

$$x^{h}_{Health,m} = \sum_{i=1}^{I} spending\ on\ healthcare\ category^{h}_{i,m}$$

vi. Adjustments to consumption aggregate

Once the components have been calculated and added up, there are two important corrections that need to be applied: 1) adjustment for regional and temporal price differences; and 2) adjustment for household composition.

(a) Price adjustment across regions and months

Nominal expenditures are affected by substantial price differences between urban and rural areas, and between different parts of the country. In addition, there are temporal price differences whereby prices faced by households at the beginning of the survey might differ from prices at the end of the survey timeline. To ensure comparability of the aggregate across geographical areas and regional price patterns, a deflator needs to be calculated and applied.

Since regional price indices were not available, this adjustment was undertaken using information on the budget shares of food items as well as the implicit prices or unit values of food items, obtained directly from the survey.
The price deflator was obtained by constructing a Paasche price index, while alternative was Laspeyres index. The difference between the two is how the relative prices are weighted. In the case of the former, the relative prices are weighted by the budget share of the item for the household itself, i.e., if a household spends a large share of its total budget on a certain food item, it receives a higher weight. In contrast, the Laspeyres formulation weights the relative prices according to the budget share of that item for a hypothetical, average reference household. In this case, all prices faced by households, irrespective of their consumption of an item, receive the same weight as that of the reference household.

The implications of the choice of price index are important, if there are significant differences in relative prices faced across regions within a country and if the hypothetical reference household is not representative of significant parts of the country. If this is the case, Paasche allows the most flexibility to incorporate these differences in relative prices and budget shares. In addition, construction of Laspeyres requires that composition of the price basket is largely similar across regions, but this is rarely the case as regions consume differential diets. Based on this it was decided to use the Paasche formulation for the spatial and temporal price index. Unlike other candidate measures, Paasche incorporates the quantities of each item consumed by the household in a specific month and weights them accordingly.

The Paasche price index for each household is obtained with the following formula:

$$P_m^h = \frac{\sum_i uv_{i,m}^h * quantity_{i,m}^h}{\sum_i uv_i^0 * quantity_i^h}$$

Where uv_i^0 is the price of food item *i* for the reference strata 0, which is in our case a median for total population, covering the survey period, i.e. at country level.
An intuitive interpretation of the Paasche index is that it represents the ratio between the cost of the food basket spent by a household, in a specific month over the cost of the same basket for the average, national, hypothetical household, here indexed by "0". The Paasche index can also be presented in another way (Deaton & Zaidi, 2002):

$$P_m^h = \left(\sum_{i=1}^{I} w_{i,m}^h * \frac{uv_i^0}{uv_{i.m}^h} \right)^{-1}$$

Where:
$w_{i,m}^h$ is the budget share of food item *i* in food basket for a given household in a specific month
$\frac{uv_{i.m}^0}{uv_i^h}$ is the relative price of the item *i*.

The Paasche index was first calculated at the level of each household. Median national prices for each item, calculated by applying household weights, were used as reference prices. Then, to stabilize the index, the median value of household price indices was taken at state level. A strata/state level index was preferred to a household Paasche index, to avoid the effect of influential outliers in some households, as well as of cases in which the household food consumption basket heavily concentrated on a few food items or most of its budget spent eating outside of home.

$$PI_m^{state} = meadin(P_m^h), over\ state/strata$$

(b) Adjustment for household composition

Since, the purpose of consumption aggregate is to capture the welfare measure at the individual rather than at the household level, the total expenditure is adjusted by household composition based on household demographic variables. The resulting welfare measure can then be attributed to each household member rather than to the household unit. To this end, the welfare aggregate in addition to spatial/temporal deflation is adjusted using household size, i.e. transformed into per-capita terms, which assumes that all individuals in the household have the same needs and that consumption is shared equally among household members. An alternative is to calculate and apply adult equivalent household size, whereby an attempt is made to capture the differences in consumption practices/requirements by age and sex. The key problem with adult equivalent scale is its arbitrariness, related to the difficulty of choosing an appropriate scale. In contrast, using the household size is more appealing due to its simplicity and transparency of assumptions.

$$size_m^h = \sum_{i=1}^{I} memeber\ of\ a\ household_{i,m}^h$$

vii. Putting all components together

Once all components of aggregate are added up for each household, the resulting nominal consumption aggregate needs to be adjusted for household size and spatial and inter-temporal price differences, observed across regions and months of survey:

$$Consagg_{adj}^{pc} = \frac{x_{Food,m}^h + x_{Meals,m}^h + x_{Non-food,m}^h + x_{Education,m}^h + x_{Health,m}^h + x_{Housing,m}^h}{PI_m^{state} * size_m^h}$$

Where:

$Consagg_{adj}^{pc}$ is spatially and temporally adjusted consumption aggregate in per capita terms

$x_{Food,m}^h$ is food consumption expenditures for specific household in a specific month;

$x_{Meals,m}^h$ is expenditures on food consumed outside of home for specific household in specific month;

$x^h_{Non-food,m}$ is non- food expenditures for specific household in a specific month;
$x^h_{Education,m}$ is expenditures on education for specific household in a specific month;
$x^h_{Health,m}$ is expenditures on health care for specific household in a specific month;
$x^h_{Housing,m}$ is housing expenditures, rent for specific household in a specific month;
PI^{state}_m is spatial and temporal price deflator, for each state in a specific month;
$size^h_m$ is the household size.

In accordance with the general principles of computing and adjusting the consumption aggregate outlined above, the structure of the GHS-Panel 2018-19 calls for additional adjustment. Data collection of the GHS-Panel 2018-19 was carried out in two visits. Visit 1 data were collected during the months of July and August of 2018- designated as "post-planting" agricultural season. Visit 2 data were collected during the months of January and February of the following calendar year – "post-harvest" agricultural season. Some modules, relevant for construction of the aggregate were not administered in both visits. For example, modules for select non-food, education and health expenditures were collected in visit 2, but not in visit 1, while the rent and housing module on which the housing rent is imputed is available in visit 1 only.

Given the panel structure of data and its seasonal pattern, it was decided to average the values of each component/group over two visits. This would allow for smoothing the seasonal variations in the consumption expenditures (primarily in food). This also implies that for the components that are only available in one visit and not the other (e.g. education, health, select non-food, rent) the values remain constant at the level of the visit when the subject module was administered.

In the final step the spatially adjusted consumption aggregate is normalized to keep the same mean as in nominal distribution. This will not alter the statistical properties / moments of distribution but has an advantage of preserving the mean and scale of distribution.

viii.Data file

The estimates of consumption aggregate as per procedure outlined above is presented in the Stata file "***totcons_final.dta***". The file contains the following set of variables:

Variable	Description	Note
hhid	Household identifier	
sector	Rural-urban location	Geographic identifier, visit 2
zone	Geo zone location	
ea	Enumeration area	
hhsize	Total number of household members	Visit 2
wt_wave4	Sampling weight (household, wave 4)	Statistical weights, visit 2
popw	Population weight (individual = wt_wave4 * hhsize)	
food_own#	Consumption expenditures for food group # from self-production or received as a gift	Annual expenditures by category (average of two visits, in per capita terms)
food_purch#	Consumption expenditures for food group #, purchased	
food_meals20	Expenditures on food consumed outside of home	
nonfood21	Expenditures on transport, other fares	
nonfood22	Expenditures on fuel	
nonfood23	Expenditures on electricity	
nonfood24	Expenditures on entertainment	
nonfood25	Expenditures on personal services	
nonfood26	Expenditures on household nonfood items	
nonfood27	Expenditures on clothing, footwear	
nonfood28	Expenditures on petrol	
nonfood29	Expenditures on carpets, mattress	
edu29	Expenditures on school tuition	
edu30	Expenditures on other education related goods and services	
health31	Expenditures on healthcare consultations and drugs	
health32	Expenditures on hospitalization	
rent33	Rent/housing expenditures (imputed for owners and actual for renters)	
reg_def_mean	Spatial and temporal price deflator	Average of two visits
totcons_pc	Total consumption expenditures, annual, nominal, in per capita terms	Average of two visits
totcons_adj	Total consumption expenditures, annual, adjusted for spatial deflation, in per capita terms	Average of two visits
totcons_adj_norm	Total consumption expenditures, annual, adjusted for spatial deflation, normalized preserving the nominal mean, in per capita terms	Final welfare aggregate (average of two visits)

Appendix 5: Changes to the Data

February 2021

Geospatial variables

- Two data files (***NGA_HouseholdGeovars_Y4.*dta** and ***NGA_PlotGeovariables_Y4.dta***) were added which contain geospatial variables derived from household and plot coordinates. For more information on what is contained in these files, the methodology, and the source of geospatial data, see Appendix 3.

Post Harvest - Agriculture

Section A3ii: Crop Disposition

- Description: Information on the total quantity harvested of the crop (Q1) was collected only if the harvest quantity was reported in different unit/size/condition across plots in Section A3i. For those cases where the unit/size/condition was consistent across plots in Section A3i, the CAPI program would automatically calculate the total harvest quantity of the crop across all household plots. This automatic sum was previously not included in the data files but has now been populated into the responses to Q1 (quantity, unit, size, condition, and conversion). The unit/size/condition reported in Q1 corresponds to the unit/size/condition that applies to the quantities of crop sold, consumed, etc. in Q5a and Q11aa-Q18a.
- Relevant file:
  - secta3ii_harvestw4
- Variables affected:
  - sa3iiq1a
  - sa3iiq1c
  - sa3iiq1d
  - sa3iiq1b
  - sa3iiq1_conv

October 2021

Consumption aggregates

- The data file (***totalcons_final.dta***) containing the calculated household consumption expenditure aggregates were added to the microdata files. Documentation on the calculation of the consumption aggregates as well as the variables contained in the data file can be found in Appendix 4 above.

Long-panel weights

- Weights for the long-panel sample (*wt_longpanel*) were added to both the cover page data files (***secta_plantingw4.dta*** and ***secta_harvestw4.dta***). These weights are suitable for panel analysis (going back to wave 1) for the long-panel sample. Further documentation on the calculation of the long-panel weights was also added in Section 3.2 above.