THE FEDERAL REPUBLIC
OF NIGERIA

NATIONAL BUREAU OF STATISTICS

LSMS
Living Standards Measurement Study

Basic Information Document

Nigeria
General Household Survey–Panel
2015/16

December 4, 2016

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 panel survey component of the General Household Survey (GHS) fielded by the National Bureau of Statistics (NBS) in 2015-2016. This survey is the third wave of a panel survey of households. The GHS-Panel is the result of a partnership that the NBS has established with the Federal Ministry of Agriculture and Rural Development (FMA&RD), the National Food Reserve Agency (NFRA), 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 new and 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 GHS survey is a cross-sectional survey of 22,000 households carried out periodically throughout the country. 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 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 third wave of the GHS-Panel was carried out in two visits (post-planting visit in September – November 2015 and post-harvest visit in February-April 2016).

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 the Wave 2 GHS-Panel. Wave 2 consisted of two visits to the 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 agriculture 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 and also administered a household and a community questionnaire.

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

2.0 The Survey Instruments

The GHS-Panel Wave 3 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 and child immunization); labour and labour data collection options; 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 (See Section 7.2.6 and Appendix 4). The labour module of the Household Questionnaire introduced four different variants to test the sensitivity of labour statistics to how labour modules are designed.

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 7.2.2 for more details.

The Agriculture Questionnaire collects different information during each visit, but for the same plots and crops. See Section 7.2.3 for more details.

The Community Questionnaire collected prices during both visits, and different community level information during the two visits. See Section 7.2.4 for more details.

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

[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).

Table 2.1: GHS-Panel Household Wave 3 Questionnaire – Post Planting Visit

Section	Topic	Respondent	Description
Cover	Cover	Field staff	Household identifier variables, enumerator, supervisor, and data entry clerk identifiers, date and time of interview and data entry, and observation notes by enumerator regarding the interview
1	Roster	Household head or spouse.	Roster of individuals living in the household, relationship to the household, gender, year of birth, age, marital status, spouse identification, parental status, and place of birth.
3	Labour	Individuals 5 years and above	Labour market participation during the last seven days, wage work, and domestic activities within the home
4A	Savings and Insurance	Individuals 15 years and above	Savings made, insurance, and remittances by the household during the last six months, and conditions of the transaction
4B	ICT – Mobile Phone Banking	Individuals 10 years and above	Data on mobile phone access and usage habits
4C	Credit	Individuals 15 years and above	Data on credit history including loans received, loans pending, or loan refusals
5	Household assets	Household head	Ownership of assets and value
7A	Meals Away From Home	Most knowledgeable person	Naira value of food consumed outside the home during the last seven days.
7B	Household Food Expenditure	Person responsible for food purchases	Quantity and value of food consumed within the household during the last seven days.
8	Household Non-food Expenditures	Person responsible for household purchases	Non-food expenditure during the last week/last month/last six months/last 12 months
9	Food Security	Household head or eligible adult	Food security status of households in during the past 7 days/12 months
9B	Subjective wellbeing	Senior female or person most knowledge about food consumption	Data on the stability of food consumption required to meet nutritional needs
11	Housing		Data on homeownership and characteristics of home (e.g. type of roof, floor, number rooms, heating source)

Table 2.2: GHS-Panel Agriculture Wave 3 Questionnaire – Post Planting Visit

Section	Topic	Respondent	Description
Cover	Cover	To be completed by field Staff. Household ID must be copied from Household to Agriculture Questionnaire.	This section contains household location and identification data as well as administrative data as regards administering and managing the questionnaire
11A	Plot Roster	Owner or manager of plot	Information on all plots owned and/or managed by the Household. This section includes data on estimated area, GPS measured area and the GPS measured location of the plot
11B1	Land Inventory	Owner or manager of plot	Data on plot acquisition, tenure and use
11C1	Planting Labour	Owner or manager of plot	Information on household members and hired labour that worked in planting activities on the plot. Includes information on amount of time spent by each person and payments made to hired labour
11E	Seed acquisition	Owner or manager of plot	Data on source, quantity and cost of seeds used on the plot
11F	Planted field crops	Owner or manager of plot	Data on crops planted on the plot, amount of crops planted and expected harvest. Also includes questions on cowpeas, variety, features and year of adoption
11I	Animal holdings	Farmer or caretaker of animals	Data on farm animals owned by the household and commercial activity with these animals
11J	Animal costs	Farmer or caretaker of animals	Livestock farmer caretaker activities and costs
11K	Agriculture by-product	Farmer or caretaker of animals	Trading activity in agricultural by-products
11L1	Extension Services I	Owner or manager of plot	Main source (government and non-governmental) of farming advice on select agricultural activities
11L2	Extension Services II	Owner or manager of plot	Details of information provided by main source of information on agricultural activities. Includes frequency of visits and usefulness of the information provided
12	Network Roster	Farmer, owner or manager of plot	Roster of places or businesses where the household sells and purchases agricultural produce and/or supplies

Table 2.3: GHS-Panel Community Wave 3 Questionnaire – Post Planting Visit

Section	Topic	Respondent	Description
Cover	Cover	To be completed by the field staff	Cover
C1	Respondents Characteristics	Community Focus Group	Respondent characteristics
C2	Food Prices	Market Food Sellers	Food prices
C3	Labour	Community Focus Group	Labour
C4	Land Prices and Credit	Community Focus Group	Land prices and credit

Table 2.4: GHS-Panel Household Wave 3 Questionnaire – Post Harvest Visit

Section	Topic	Respondent	Description
Cover	Cover	To be completed by the field staff	Household identifier variables, enumerator, supervisor, and data entry clerk identifiers, date and time of interview and data entry, and observation notes by enumerator regarding the interview
1	Roster	Head of Household or spouse.	Roster of individuals living in the household, relationship to the household, gender, year of birth, age, marital status, polygamous marriages, spouse identification, parental status, place of birth, date joined household if new, migration.
2	Education	Individuals 5 years and above	Educational attainment, school characteristics, and expenditures.
3	Labour	Individuals 5 years and older	Labour market activity and information on employment in one or more industries in the past 6 months of all household members 5 years and older. This includes employment and earnings information.
4	Health	All individuals	General health status, utilization and cost of health services for those that need medical care. Data on effect of disabilities on activity and functioning; and anthropometrics. Child immunization records.
4B	Child Development	Mother or primary caretaker of each child between 2 and 18 yrs	Information about ability to communicate and motor skills
6	Remittances	All individuals 10 years and above	Remittances received from abroad by household members 10 years and older

Section	Topic	Respondent	Description
6A	Behaviour	Head of household, spouse, or other senior member	Data on household preferences
6B	Attitude	Head of household or other senior member	Data on emotional or mental state
10B	Food Expenditures	Female in the household responsible for food preparation and/or food purchases	Data on the content and cost of meals prepared by the household in the past 7 days
10C	Aggregate Food Consumption	Female in the household responsible for food preparation and/or food purchases	Data on the content of meals consumed by the household in the past 7 days overall and by various age groups
11	Non-food Expenditures	Most knowledgeable person or person who is responsible for household purchases	Consumption and expenditure on non-food items
12	Food Security	Household head or knowledgeable adult household member	Collects information on quantity of food, preferred foods and variety of foods available to household members based on economic reasons. Also collects data on intra-household food security dynamics.
13	Other household Income	Household head or knowledgeable adult household member	Miscellaneous income received by household
14	Safety Nets	Household head or knowledgeable adult household member	Household access to and utilization of safety nets
15A	Economic Shocks	Household head or knowledgeable adult household member	Data on economic shocks affecting the household
15B	Deaths	Household head or knowledgeable adult household member	Deaths of household members in the past 12 months, including age of deceased and cause of death.
15C	Conflict		Data on the frequency of various types of violence experienced by any member of the household

Table 2.5: GHS-Panel Agriculture Wave 3 Questionnaire – Post Harvest Visit

Section	Topic	Respondent	Description
Cover	Cover	To be completed by field Staff. Household ID must be copied from Household to Agriculture Questionnaire.	This section contains household location and identification data as well as administrative data as regards administering and managing the questionnaire
A1	Land	Farmer, owner or manager of plot	Follow-up on use of land for in post-planting visit and data on any subsequent planting or other use of the plot. Also information collected on new plots (i.e. added since post-planting visit)
A2	Labour	Farmer, owner or manager of plot	Information on household members and hired labour that worked in crop harvesting activities on the plot. Includes information on amount of time spent by each person and payments made to hired labour
11C2	Input Cost		Information about pesticide use, herbicide use, animal traction, equipment and machinery use
11D	Fertilizer Acquisition		Left over (inorganic fertilizer), free (inorganic fertilizer), source of commercial (inorganic fertilizer), organic fertilizer use
A3	Agricultural production/ disposition	Farmer, owner or manager of plot	Quantity and value of field crops produced
A4	Agricultural Capital	Farmer, owner or manager of plot	Ownership and value of agricultural machinery and tools owned by the household
A5	Extension Services	Farmer, owner or manager of plot	Access to and utilization of technical support from various sources (government and non-government)
A8	Other Agricultural Income	Farmer or caretaker of animals	Income from sale of agricultural products not capture previous section under crops and livestock
A9	Fishing	Owner of fishing operations	Data on fishing activities, includes capture, harvesting and processing. Information on boat usage and the use of hired labour
A10	Network Roster	Farmer, owner or manager of plot	Roster of places or businesses where the household sells and purchases agricultural produce and/or supplies

Table 2.6 GHS-Panel Community Wave 3 Questionnaire – Post-Harvest Visit

Section	Topic	Respondent	Description
Cover	Cover	To be completed by the field staff	Cover
C1	Respondents Characteristics	Community Focus Group	Respondents Characteristics
C2	Community Infrastructure and Transportation	Community Focus Group	Community Infrastructure and Transportation
C3	Community Organizations	Community Focus Group	Community Organizations
C4	Community Resource Management	Community Focus Group	Community Resource Management
C5	Community Changes	Community Focus Group	Community Changes
C6	Community Key Events	Community Focus Group	Community Key Events
C6A	Conflict	Community Focus Group	Data on violent events occurring in the community
C7	Community Needs, Actions, and Achievements	Community Focus Group	Community Needs, Actions, and Achievements
C8	Food Prices	Market Food Sellers	Food Prices

There were some changes made in the questionnaires between Waves 2 and 3 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). Tables 2.7 and 2.8 outline these changes for the post-planting and post-harvest visits, consecutively.

Table 2.7: Wave 2 to Wave 3 Comparison, Post-Planting

Questionnaire	Section	Notes
Household Questionnaire	Section 1: Household Roster	Questions re-ordered in Wave 3: starting with Q7
		Questions added in Wave 3: Q13b, Q13c
		Question dropped in Wave 3: Q15, Q33, A35-42
	Section 2: Education	Section dropped
	Section 3A: Labour	Wave 3: consolidated into Section 3
		Questions added in Wave 3: Q5b, Q5c, Q6b, Q6c, Q12b, Q15b-Q15f, Q28b-Q28f,
		Questions dropped in Wave 3: Q12, Q19, Q20, Q32, Q33,

Questionnaire	Section	Notes
	Section 3B: Labour 6 Months, 12 Months, Activity Table, and Activity Summary	Wave 3: consolidated into Section 3
	Section 4: Credit and Savings	Appear as two sections in Wave 3: Section 4a – Savings and Insurance and Section 4c - Credit
		Section 4a – Savings and Insurance, Change in Wave 3: Q11-Q15 replaced
	Section 4B: Financial Capability	Removed in Wave 3
	Section 4B: ICT – Mobile Phone Banking	Added in Wave 3
	Section 5: Household Assets	No changes
	Section 6: Nonfarm Enterprises and Income Generating Activities	Removed from Wave 3 (information is only collected in post-harvest Section 9).
	Section 7A: Meals Away from Home	No changes
	Section 7B: Food Expenditure	Change in Wave 3: Changes to food and unit codes
	Section 8: Non-Food Expenditure	Change in Wave 3: Changes to recall codes
	Section 9: Food Security	No changes
	Section 9B: Subjective Wellbeing	New section in Wave 3
	Section 10: Other household income	Removed in Wave 3
	Section 11: Housing	New section in Wave 3
Agriculture Questionnaire	Section 11A: Plot Roster	Questions added in Wave 3: Q4a, Q4b
	Section 11B1: Land Inventory	Question re-worded or options modified in Wave 3: Q4, Q9
		Questions added in Wave 3: Q8a, Q8b, Q9a, Q16b, Q19a, Q28a, Q47-Q51
	Section 11B2: Land Tenure	Dropped in Wave 3
	Section 11C1: Planting Labour	Section Renamed – “Labour” in Wave 3
		Addition directions to reference Section 11B1, Question 27 for cultivated plots in Wave 3

Questionnaire	Section	Notes
		Questions added in Wave 3: Q11 and Q12
	Section 11C2: Input Cost	Moved to Wave 3, Post Harvest
	Section 11D: Fertilizer Application	Moved to Wave 3, Post Harvest
	Section 11E: Seed Acquisition	Questions added in Wave 3: Q3a, Q3b
	Section 11F: Planted Field Crops	Section renamed Planted Field and Tree Crops
		Question modified in Wave 3: Q4
		Q4 is Q12 in Wave 3
		Questions added in Wave 3: Q13 – Q15
		Questions dropped in Wave 3: Q5 – Q12
	Section 11G: Planted Tree Crops	Combined with 11F in Wave 3, starting at Q4
	Section 11H: Marketing	Dropped in Wave 3
	Section 11I: Animal Holdings	Questions dropped in Wave 3: Q1, Q2
		Condition "in new year" dropped in Wave 3: Q9 – Q21
	Section 11J: Animal Costs	No change
	Section 11K: Agriculture By-Product	Condition "in new year" dropped in Wave 3: Q1
	Section 11L1: Extension Services I	No change
	Section 11L2: Extension Services II	Referred to as Extension Services 2 in Wave 3
	Section 12: Network Roster	No change
Community Questionnaire	Section 1: Respondent Characteristics	No change
	Section 2: Food Prices	Change in Wave 3: additional item codes differ
		Q1 starts at Q7 in Wave 3
		Q1 – Q6 collect geographical information about the main sources of food (i.e. marketplace, shops, or other)
	Section 3: Labour	Q14 – Q23 dropped in Wave 3
		Wave 3 include additional activity codes
	Section 4: Land Prices and Credit	No change

Table 2.8: Wave 1 to Wave 2 Comparison, Post-Harvest

Questionnaire	Section	Notes
	Cover	Wave 3 adds: AG1 – AG3
	Section 1: Household Roster	Change in Wave 3: adds Q5, Q12a-Q12c
		Questions dropped in Wave 3: Q10
		Q14 moved to Q4a
	Section 2A: Education – New Members	Appears as Section 2 in Wave 3
		Wave 3 drops Q11, Q12, Q28
	Section 2B: Education – Original Household Members	Wave 3 consolidates 2A and 2B into single Section 2. Education only collected in post-harvest.
	Section 2: Education	Single education section in Wave 3
	Section 3A: Labour	Wave 3 consolidates 3A and 3B into single Section 3
		Q1 appears as Q4 in Wave 3, subsequent questions are re-ordered thereafter
		Questions dropped in Wave 3: Q12, Q16, Q17, Q24b, Q28, Q29, Q33, Q34, Q36-38
		Questions added in Wave 3: Q28b, Q28c, Q28d, Q28e, Q28f
		Q30b appears as Q35 in Wave 3
		Q39 time code changed
	Section 3B: Labour – 12 Months	Wave 3 consolidates 3A and 3B into single Section 3
		Begins at Q42 in Wave 3, time codes are reformatted
	Section 3: Labour	New to section Wave 3, consolidates 3A and 3B from Wave 2
		Adds new Q1 confirming if the household member if 5 years or older
	Section 4A: Health	Added to Wave 3: Q37a
		In Wave 3 Q51 modified to: “a child aged less than 84 months (less than 7 years)”
		Dropped in Wave 3: Q22e, Q24-Q34, Q36, Q40-Q50
	Section 4B: Child Immunization	Not included in Wave 3
	Section 4: Health	Section 4A as specified above is included in this section
	Section 4B: Child Development	New to Wave 3
	Section 5: Information and Communication Technology	Not included in Wave 3
	Section 6: Remittances	Dropped in Wave 3: Q3, Q7

Questionnaire	Section	Notes
	Section 6A: Behavior	New to Wave 3
	Section 6B: Attitude	New to Wave 3
	Section 7: Household Assets	Not included in Wave 3 (Only collected in post-planting)
	Section 8: Housing	Now collected in post-planting Section 11
	Section 9: Nonfarm Enterprises and Income Generating Activities	Change in Wave 3: Added - Q4a, Q13, Q27a; New response options - Q10, Q13
	Section 10A: Meals Away From Home	No Change
	Section 10B: Food Expenditure	Change in Wave 3: modified food item and unit codes
	Section 10C: Aggregate Food Consumption	No change
	Section 11: Non-food expenditure	Change in Wave 3 item options (e.g. one-month recall)
	Section 12: Food Security	No change
	Section 13: Other Household Income	Q6 in Wave 3 includes additional condition of "past 12 months"
	Section 14: Safety Nets	Wave 3 includes an additional item codes
	Section 15A: Economic Shocks	No change
	Section 15B: Deaths	No change
	Section 15C: Conflicts	New to Wave 3
Agriculture Questionnaire	Section A1: Land and Dry Season Planting	Section renamed "Land" in Wave 3
		Added to Wave 3: Q4b, Q10a – Q10c
		Dropped in Wave 3: Q26 – Q28
	Section A2: Harvest Labour	Renamed "Labour" in Wave 3
		New component – Hired Labour between planting and harvesting using same convention as other sections, expect: Q1 split into Q1a and Q1b in Wave 3, Q2 – Q4 appear as Q1c – Q1e in Wave 3 (respectively), Q5 – Q7 appear as Q1f – Q1h in Wave 3 (respectively), Q8 – Q10 appear as Q1i – Q1k in Wave 3 (respectively)
		Question reassigned to the following number in Wave 3: Q10b to Q1l, Q11 to Q1m, Q12 to Q1n, Q13 to Q1o
		Added to Wave 3 – Household Labour (For Harvesting and Threshing)

Questionnaire	Section	Notes
		Added to Wave 3 – Hired Labour (For Harvesting and Threshing)
	Section 11C2: Input Cost	Moved from post-planting in Wave 3
	Section 11D: Fertilizer Acquisition	Moved from post-planting in Wave 3
	Section A3i: Agricultural Production – Harvest of Field and Tree Crops	Assigned subsection A3i in Wave 3 and renamed "Crop Harvest"
		Added in Wave 3: Q4, Q4a, Q6, Q6a – Q6e
		Number change to in Wave 3: Q6B to Q6e
		Dropped in Wave 3: (modified) versions of Q7 appear in A3ii
	Section A3ii: Agricultural Production – Crop Disposition	New to Wave 3
	Section A4: Agricultural Capital	Added in Wave 3: Q3b
	Section A6: Animal Holdings	Moved to Post-planting in Wave 3
	Section A7: Animal Cost	Moved to Post-planting in Wave 3
	Section A5a: Extension Services (Topics)	No changes
	Section A5b: Extension Services (Sources)	No changes
	Section A8: Other Agricultural Income	Renamed Other Agricultural Income: By Product in Wave 3
	Section A9a: Fishing	New unit codes added in Wave 3 for Q4, Q5, Q7, Q10, Q12, Q13, Q15, Q18
	Section A9b: Fishing Capital and Revenues	Dropped in Wave 3: Q23 – Q25
	Section A10: Network Roster	No changes
Community Questionnaire	Section C1: Respondent Characteristics	No change
	Section C2: Community Infrastructure and Transportation	Q8 is modified and appears as Q2 in Wave 3
		Q7 is dropped in Wave 3
		Additional item codes in Wave 3

Questionnaire	Section	Notes
	Section C3: Community Organizations	No change
	Section C4: Community Resource Management	No change
	Section C5: Community Changes	Wave 3 is compared to 5 years ago
	Section C6: Community Key Events	No change
	Section C6A: Conflict	New in Wave 3
	Section 7: Community Needs, Actions and Achievements	Q1 Wave 3 reflects actions in the past 5 years
	Section C8: Food Prices	Divided into two sections in Wave 3 - C8a and C8b
		Q1a, Q1b and Q2 appears at Q1, Q2, and Q3 respectively in Wave 3
	Section C8a: Food Prices 1st Location	Geolocation information captured by Q1 – Q6 in Wave 3
	Section C8b: Food Prices 2nd Location	Geolocation information captured by Q1 – Q6 in Wave 3

3.0 Wave 3 Sample and Weights

The GHS-Panel sample is designed to be representative at the national level as well as at the zonal (both urban and rural) level. The sample size of the GHS-Panel is not adequate for state-level estimates. The complete sampling information for the GHS-Panel Wave is described in the Basic Information Document for GHS-Panel 2010/2011.

The objective of the GHS-Panel Wave 3 was to re-interview all of the Wave 2 households. An effort was also made to locate and interview households that were not interviewed during Wave 2. 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.

Table 3.1 shows the details of the Wave 3 sample. The Wave 3 sample size for households interviewed in both post-planting and post-harvest visit is 4581. This size is only 135 households less than Wave 2. However, there were some households that were not interviewed in Wave 2 that were found and interviewed in Wave 3.

Table 3.1: Details of GHS-Panel Sample in each Wave

	Wave 1			**Wave 2**			**Wave 3**		
	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural
North Central	794	217	577	784	214	570	777	210	567
North East	797	138	659	741	117	624	643	106	537
North West	898	170	728	878	156	722	882	163	719
South East	794	204	590	763	197	566	755	193	562
South South	769	229	540	761	219	542	744	221	523
South West	864	611	253	789	562	227	780	556	224
Total	4916	1569	3347	4716	1465	3251	4581	1449	3132

4916 is the number of households that are in Wave 1 post-planting **and** Wave 1 post-harvest
4716 is the number of households that are in Wave 2 post-planting **and** Wave 2 post-harvest
4581 is the number of households that are in Wave 3 post-planting **and** Wave 3 post-harvest

Table 3.2 shows the size for three different samples. The first is the sample of households interviewed in both visits for all three waves. The second is the sample of households interviewed in both visits of Wave 1 and Wave 3. The third is those interviewed in both visits of Wave 2 and Wave 3. The full, balanced panel (across all three waves) consists of 4,407 households. For the "long" panel (Wave 1 and 3) there are 4,533 households and for the "short" panel there are 4,448 households.

Table 3.2: GHS-Panel Sample for Combinations of Waves

	Wave 1, Wave 2, & Wave 3			**Wave 1 & Wave 3**			**Wave 2 & Wave 3**		
	All	Urban	Rural	All	Urban	Rural	All	Urban	Rural

North Central	767	206	561	772	207	565	772	209	563
North East	592	85	507	641	104	537	593	86	507
North West	868	153	715	880	163	717	870	153	717
South East	738	186	552	752	190	562	740	188	552
South South	710	204	506	722	213	509	730	211	519
South West	732	521	211	766	546	220	743	529	214
Total	4407	1355	3052	4533	1423	3110	4448	1376	3072

Since the GHS-Panel is a panel survey, every effort is made to maintain as many households as possible in the sample. However, there is also some attrition for a variety of reasons. The left side of Table 3.3 shows the attrition between Wave 1 and Wave 2 as well as between Wave 2 and Wave 3. Compared to Wave 1 to 2, attrition was higher between Wave 2 and 3 by 23 households. However, a large majority of these households were lost as a result of the poor security situation in the North East Zone. More than half of the attrition was concentrated in this Zone. As a result of the security situation, 14 EAs could not be visited in Borno and Yobe States. As shown in Table 3.4, those 14 EAs contained 137 households that could not be interviewed. Other reasons for attrition are also contained in Table 3.4. Fifty-seven households could not be located or any information about their current location acquired. An additional 25 households refused to continue to participate in the survey.

Table 3.3: Sample Attrition between Waves

	Attrition					
	Wave 1 to Wave 2			**Wave 2 to Wave 3**		
	All	Urban	Rural	All	Urban	Rural
North Central	15	6	9	12	5	7
North East	57	22	35	148	31	117
North West	22	14	8	8	3	5
South East	34	10	24	23	9	14
South South	29	17	12	31	8	23
South West	88	58	30	46	33	13
Total	245	127	118	268	89	179

Table 3.4: Reason for Attrition (Wave 2 to 3)

Reason	**# of HH**
Single Household EA	1
Refused	25
Not found	57
Dead	40
Moved away (not tracked)	8
Crisis area	137

In order to retain households that had moved away, two tracking exercises were conducted following the post-planting visit and again the post-harvest visit. Table 3.5 shows the number of

households that were successful tracked in both of these visits. A total of 186 and 48 households were successfully tracked following the post-planting and post-harvest visits, respectively. Tracking was more substantial following post-planting since households that had moved since Wave 2 (over two years prior) were being tracked whereas for post-harvest only households that had moved between post-planting and post-harvest of Wave 3 were being tracked. As for Wave 2, the tracking was largely concentrated in the southern Zones, especially South West.

Table 3.2: Distribution of households without complete panel information

	Wave 3 Tracking					
	Post-Planting			**Post-Harvest**		
	All	Urban	Rural	All	Urban	Rural
North Central	10	4	6	5	1	4
North East	15	4	11	6	3	3
North West	9	3	6	1	1	0
South East	25	13	12	3	1	2
South South	37	13	24	12	3	9
South West	90	68	22	21	13	8
Total	186	105	81	48	22	26

When a sample of households is selected for a survey, these households represent the entire population of the country. To accurately use the datasets, the data must be weighted to reflect the distribution of the full population in the country and the original sampling frame. The main weighting procedure is described in more detail in the Wave 1 Basic Information Document. In Wave 3, we also provide a detailed description in Appendix 5 of this document. The original wave 1 weights were adjusted to account for attrition. However, the survey weights should reflect the sample used for the analysis. In order to provide a more comprehensive set of weights, we provide different survey weights for 6 different samples in Wave 3. Table 3.6 lists the relevant samples, the variable name for the survey weights, and the location of the weights. In order to prevent confusion, only the Wave 3 (*wt_wave3*) and full panel (*wt_w1w2w3*) weights are included in the household cover sheet files (***secta_plantingw3.dta*** and ***secta_harvestw3.dta***). The full set of available weights (including from Wave 1 and Wave 2) are included in ***HHTrack.dta***.

Table 3.6: Survey Weight Samples, Variable Names, and Locations

Sample	Variable name	Location (data file)	
		HHTrack	*Cover sheets (secta_…)*
Wave 3, post-planting visit	*wt_w3v1*	X	
Wave 3, post-harvest visit	*wt_w3v2*	X	
Wave 3, both visits	*wt_wave3*	X	X
Wave 1, 2 &, 3*	*wt_w1w2w3*	X	X
Wave 1 & 3*	*wt_w1w3*	X	
Wave 2 & 3*	*wt_w2w3*	X	

*Interviewed in every visit

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 team of the survey participated in the TOT, which lasted for five days. 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 was conducted over a nine-day period. Seven days were dedicated to theory including data entry training and two days to field practice and review. Participants in the training were Zonal Controllers, State Officers, Field Supervisors, Field Interviewers, and Data Entry Operators. Training instructions were given to the field staff by the resource persons from the management team (NBS, FMS&RD, and NFRA) with support from World Bank technical missions. Three (3) 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 and (iv) data entry 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

Due to security concerns in the North-East and North-West zones, the training for those two zones was moved to North-Central zone. As a result, the North-East and North-West were trained in two training locations near the town of Karu in Nasarawa State. The training for the South-East, and South-South zones was conducted in Enugu (Enugu State). The training for the South-West and North-Central zones was conducted in Ibadan (Oyo State).

4.4 Evaluation of Field and Data Entry Staff

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 entry operators were removed from the survey. In some instances, the removed workers were replaced and in other cases there was no replacement but those remaining in the team were given extra time to complete the fieldwork and data entry.

5.0 Field Work

5.1 Organization of Fieldwork

Data were collected by teams consisting of a supervisor, between 2 and 4 interviewers, and a data entry operator. 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 2 was administered in two visits: post-planting (September - November 2015) and post-harvest (February - April 2016). A tracking phase was conducted after both visits in October-November 2015 and April-May 2016 to interview households that had moved from their location in the previous visit (Wave 1 or Wave 2) or had moved between Visit 1 and Visit 2 in Wave 3.

5.2 Gift to Households

As a show of appreciation for the panel households continued participation, all panel households that were located, were given a gift (even if they refused to participate). These gifts were given during the post-planting survey and consisted of either a torchlight or a rechargeable lamp. 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-printed Information

To facilitate identification of the same people over time, the field team implemented Wave 3 with a pre-printed household roster. The roster asks for information on all Wave 1 members (whether they still reside in the household, have moved or are deceased). New members are added to the roster. So the ID number in the roster can be merged with Wave 1 and Wave 2 to identify the same respondent.

Interviewers were also provided with prefilling forms in the second (post-harvest) visit that contained information from the first visit. This included: (1) an updated household roster, (2) plot roster, and (3) plot-crop roster. During the post-harvest visit, interviewers were also provided with prefilling form containing the list of nonfarm enterprises collected in wave 1 and wave 2. This was done to more effectively maintain a panel of nonfarm enterprises.

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 first monitoring team also included World Bank officials and consultants. 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. The third monitoring focused on data issues and included checking the entered data against data in the questionnaires. Where problems were found, these were corrected either directly or through a revisit to the household for verification of information or for further information.

5.5 Methodological Experiment

In the post-harvest visit, a methodological experiment was implemented. For this experiment, two versions of the household questionnaire were administered. The only difference between these two versions is the placement of the *Behaviour* and *Attitude* sections (*Section 6A* and *Section 6B*). In version 1 these sections were asked in the middle of the questionnaire (following *Section 6: Remittances*). In version 2, these two sections were asked at the end of the questionnaire (following *Section 15C: Conflicts*). Households *within each EA* were randomly assigned to receive one of the two versions. If an EA had 10 households, 5 were randomly assigned version 1 and 5 assigned version 2. The two different version of the questionnaire are provided on the website (***PH_W3_Household_Quest_VERSION*[X].*pdf***). In the data sets, both versions for Sections 6A and 6B have been combined into single files (***sect6a_harvestw3.dta*** and ***sect6b_harvestw3.dta***). The different version administered is indicated in both files by the variable ***version***.

6.0 Household Tracking Exercise

There were two separate tracking exercises conducted in Wave 3. 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 that moved between Wave 2 and the first visit of Wave 3 were tracked. During the post-harvest tracking exercise, households that moved between the post-planting and post-harvest visits of Wave 3 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 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
ABIA	5	YES	1	1
ADAMAWA	4	YES	1	1
AKWA IBOM	7	YES	1	1
ANAMBRA	8	YES	1	1
BAYELSA	6	YES	1	1
BENUE	5	YES	1	1
CROSS RIVER	7	YES	1	1
DELTA	13	YES	1	1
EBONYI	3	YES	1	1
EDO	3	YES	1	1
EKITI	10	YES	1	1
ENUGU	4	YES	1	1
GOMBE	1	NO	-	2
IMO	4	YES	1	1
JIGAWA	3	YES	-	2
KADUNA	2	NO	-	2
KANO	2	NO	-	2
KEBBI	1	NO	-	2
KOGI	7	YES	1	1

KWARA	12	YES	1	1
LAGOS	25	YES	2	2
NASARAWA	2	NO	-	2
NIGER	2	NO	-	2
OGUN	19	YES	2	2
ONDO	22	YES	2	2
OSUN	3	YES	1	1
OYO	23	YES	2	2
PLATEAU	1	NO	-	2
RIVERS	15	YES	1	1
TARABA	7	YES	1	1
YOBE	2	NO	-	2
FCT ABUJA	4	YES	1	1
Total	232		27	40

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

STATE WHERE HOUSEHOLD RELOCATED	NUMBER OF HOUSEHOLDS	VISITED BY HQ STAFF	Number HQ Persons	Number State Persons
AKWA IBOM	5	YES	1	1
ANAMBRA	2	YES	1	1
BAUCHI	1	NO		2
BAYELSA	1	NO		2
CROSS RIVER	7	YES	1	1
DELTA	5	YES	1	1
EDO	1	NO		2
EKITI	2	NO		2
ENUGU	1	NO		2
GOMBE	2	NO		2
KADUNA	2	NO		2
KANO	1	NO		2
KOGI	2	NO		2
KWARA	2	NO		2
LAGOS	3	NO		2
OGUN	4	YES	1	1
ONDO	8	YES	1	1
OSUN	1	NO		2
OYO	5	YES	1	1
PLATEAU	2	NO		2
RIVERS	1	NO		2
TARABA	3	YES	1	1

YOBE	1	NO		2
Total	62		8	38

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 activity. This training took place on June 20 and 21, 2013. The persons trained were to train their partner staff as well as state officers in their assigned state. A number of trainers also had responsibility to train staff from states where no head office staff were slated to visit. The tracking fieldwork commenced round about June 25, 2013 and was completed by the end of July, 2013.

6.3 Tracking Methodology

The tracking of households included the following steps:

- Discussion of the set of tracking households with the state 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 on the tracking forms. Information was also used from the contact information on the questionnaires. That is, where households could not be contacted using the information on the tracking form, 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 address and to set appointments for the interview.
- In cases where there was no useful phone information (either in the tracking form of 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.
- Completion of the required questionnaires

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 early in 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 Gift to Tracked Households

As a show of appreciation for the tracked households continued participation in the panel survey, the tracked households were presented with their choice of either a large umbrella or a raincoat.

6.5 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_plantingw3* found in the Post-Planting Household data folder. For household interviewed during the post-harvest tracking exercise, see *tracked_obs* in the data file *secta_harvestw3* found in the Post-Harvest Household data folder.

7.0 Data Management and Description of Datasets

7.1 Data Management

7.1.1 Data Entry

The household and agricultural components of the survey were conducted using concurrent data entry approach. In this method, the fieldwork and data entry were handled by each team assigned to the state. Each team consisted of a field supervisor, 2-4 interviewers and a data entry operator. Immediately after the data were collected in the field by the interviewers and supervisors (the supervisors administered the community questionnaires and collected data on prices), the questionnaires were handed over to the supervisor to be checked and documented. At the end of each day of fieldwork, the questionnaires were then passed to the data entry operator for entry. After the questionnaires were entered, the data entry operator generated an error report which reported issues including out of range values and inconsistencies in the data. The supervisor then checked the report, determined what should be corrected, and decided if the field team needed to revisit the household to obtain additional information. The benefits of this method are that it allows one to:

- Capture errors that might have been overlooked by a visual inspection only,
- Identify errors early during the field work so that if any correction required a revisit to the household, it could be done while the team was still in the EA

The CSPro software was used to design the specialized data entry program that was used for the data entry of the questionnaires.

7.1.2 CAPI

For the first time in Wave 3, a portion of the survey was collected using Computer Assisted Person Interview (CAPI) techniques. The community questionnaire was implemented in both the post-planting and post-harvest visits of Wave 3 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. Supervisors were given tablets which they used to conduct the community questionnaire interviews. The primary purpose of this limited CAPI implementation was to test the viability of Survey Solutions for use in future surveys implemented by NBS, including succeeding waves of the GHS-Panel. Overall, implementation of Survey Solutions for the community module in Wave 3 was highly successful.

7.1.3 Data Communication System

The data communication system used in Wave 3 was highly automated. Each data entry person was given a mobile modem and once they connected to the internet the system would automatically send entered data to the head office in Abuja. The data entry persons were instructed to do this every one or two days so there was a steady flow of current data from the field to the head office.

7.1.4 Double Data Entry

For the first time in wave 3, double data entry was performed for all household and agriculture questionnaires in both the post-harvest and post-planting visits. Double data entry was implemented to identify and correct data entry errors. The first data entry (FDE) was performed in the field by data entry operators assigned to each survey team. Following completion of fieldwork, all questionnaires were shipped from the State offices to NBS headquarters in Abuja. A team of data entry operators was selected to perform the second data entry (SDE) at NBS headquarters. The SDE team consisted of a mix of data entry operators that participated in the field entry and some that did not. This team entered all household and agriculture questionnaires. When re-entering these questionnaires, any differences with the first data entry are flagged by CSPro and prompt the data entry operator to confirm what they have entered is correct. This design was intended to limit any data entry errors from FDE.

7.1.5 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 using the concurrent data entry system which was, as explained above, designed to highlight many of the errors that occurred during the fieldwork. The data was reviewed by the panel management team for inconsistencies and extreme values. 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. Identified errors 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 sent from the state to the head office of NBS where the data was reviewed again.

The second and third t stages of cleaning involved a final comprehensive review of the data primarily conducted by World Bank staff in Washington, DC in consultation with the headquarters and state offices of NBS in Nigeria. The second stage of cleaning consisted of a review of the FDE and SDE data. In general, the SDE was taken to be the most correct version. However, there was a review of some differences between the two entries where the SDE was inconsistent, out of range, or missing while the FDE was not. In these cases, FDE was used for the final raw data set that was used for the third stage of cleaning.

The third stage of cleaning involved a comprehensive review of the data coming out of the second stage. Every variable was examined individually for (1) consistency with other sections and variables, (2) out of range responses, and (3) outliers. Quite often when errors were identified, the two versions of data entry were compared and switched if one of the errors was not present in one version. In cases where the error was present in both FDE and SDE, obvious errors were corrected where possible and questionnaires were checked when deemed necessary. 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 3 was administered in two visits: post-planting (September - November 2015) and post-harvest (February - April 2016). 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 2015 (post-planting) and April/May 2016 (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 Post Harvest Only Households

In Wave 3, there were 20 households that were only visited in the post-harvest visit. Nineteen of the households were located in two EAs that could not be visited in the post-planting period due to security concerns. The remaining household was not located in post-planting (including during the tracking phase), but was located in the post-harvest. For these 20 households, the full post-harvest questionnaire was implemented. In addition, they were also administered the portion of the post-planting questionnaire that is not repeated in post-harvest (e.g. household roster, food consumption, nonfood expenditure, food security, etc.). For these repeated sections, there will be no data present in the post-planting version of the data. These households are identified in both of the cover sheet data sets (*secta_plantingw3* and *secta_harvestw3*) by the variable *phonly_obs*.

7.2.2 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.

- Roster
- Labour
- Meals Away from Home
- Food Consumption and Expenditure
- Nonfood Expenditure
- Food Security
- Other Household Income
- 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
    - Subjective Wellbeing
    - Housing
- Post-harvest only
    - Education
    - Health
    - Child Development
    - Remittances
    - Behaviour
    - Attitude
    - Nonfarm Enterprise and Income Generating Activities
    - Aggregate Food Consumption
    - Other Household Income
    - Safety Nets
    - Economic Shocks
    - Deaths
    - Conflict

Tables 7.1a and 7.1b show the sections of the Household Questionnaire and the datasets that correspond to these.

Table 7.1a: Post-planting household datasets

Section	Section Name	Dataset Filename
Cover	Cover	secta_plantingw3
1	Roster	sect1_plantingw3
3	Labour	sect3_plantingw3
4A	Savings and Insurance	sect4a_plantingw3
4B	ICT – Mobile Phone Banking	sect4b_plantingw3
4C	Credit	sect4c1_plantingw3
		sect4c2_plantingw3
5	Household Assets	sect5_plantingw3
7A	Meals Away From Home	sect7a_plantingw3
7B	Household Food Expenditure	sect7b_plantingw3
8	Household Non-Food Expenditures	sect8a_plantingw3
		sect8b_plantingw3
		sect8c_plantingw3
9	Food Security	sect9_plantingw3
9B	Subjective wellbeing	sect9b_plantingw3
11	Housing	sect11_plantingw3

Table 7.1b: Post-harvest household datasets

Section	Section Name	Dataset Filename
Cover	Cover	secta_harvestw3
1	Roster	sect1_harvestw3
2	Education	sect2_harvestw3

3	Labour	sect3_harvestw3
4	Health	sect4a_harvestw3
4B	Child Development	sect4b_harvestw3
6	Remittances	sect6_harvestw3
6A	Behaviour	sect6a_harvestw3
6B	Attitude	sect6b_harvestw3
9	Non-farm Enterprises and income generating activities	sect9_harvestw3
		sect9b_harvestw3
10A	Meals Away From Home	sect10a_harvestw3
10B	Food Expenditures	sect10b_harvestw3
10C	Aggregate Food Consumption	sect10c_harvestw3
		sect10ca_harvestw3
		sect10cb_harvestw3
11	Non-food Expenditures	sect11a_harvestw3
		sect11b_harvestw3
		sect11c_harvestw3
		sect11d_harvestw3
		sect11e_harvestw3
12	Food Security	sect12_harvestw3
13	Other household Income	sect13_harvestw3
14	Safety Nets	sect14_harvestw3
15A	Economic Shocks	sect15a_harvestw3
15B	Deaths	sect15b1_harvestw3
		sect15b2_harvestw3
15C	Conflict	sect15c_harvestw3

7.2.3 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 include additional plots households may have acquired or old plots they have disposed of since the first, post-planting visit.[2] The crop codes used in the Agriculture Questionnaire are presented in Appendix 3. 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.

[2] In theory, some plots in Wave 3 can be matched to Wave 1 and 2 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 7.2a: Post-planting Agriculture datasets

Section	Section Name	Dataset Filename
Cover	Cover	secta_plantingw3
11A	Plot Roster	sect11a_plantingw3
11B1	Land Inventory	sect11b1_plantingw3
11C1	Planting Labour	sect11c1_plantingw3
11E	Seed acquisition	sect11e_plantingw3
11F	Planted field crops	sect11f_plantingw3
11I	Animal holdings	sect11i_plantingw3
11J	Animal costs	sect11j_plantingw3
11K	Agriculture by-product	sect11k_plantingw3
11L1	Extension Services I	sect11l1_plantingw3
11L2	Extension Services II	sect11l2_plantingw3
12	Network Roster	sect12_plantingw3

Table 7.2b: Post-harvest Agriculture datasets

Section	**Section Name**	**Dataset Filename**
Cover	Cover	secta_harvestw3
A1	Land	secta1_harvestw3
A2	Labour	secta2_harvestw3
11C2	Input Cost	secta11c2_harvestw3
11D	Fertilizer Acquisition	secta11d2_harvestw3
A3	Agricultural production/ diposition	secta3i_harvestw3
		secta3ii_harvestw3
A4	Agricultural Capital	secta4_harvestw3
A5	Extension Services	secta5a_harvestw3
		secta5b_harvestw3
A8	Other Agricultural Income	secta8_harvestw3
A9	Fishing	secta9a1_harvestw3
		secta9a2_harvestw3
		secta9b1_harvestw3
		secta9b2_harvestw3
		secta9b3_harvestw3
A10	Network Roster	secta10_harvestw3

7.2.4 Community Data

Tables 7.3a and 7.3b show the sections of the community questionnaire and their corresponding data sets.

Table 7.3a: Post-planting Community datasets

Section	**Section Name**	**Dataset Filename**

Cover	Cover	sectc_plantingw3
C1	Respondents Characteristics	sectc1_plantingw3
C2	Food Prices	sectc2_plantingw3
C3	Labour	sectc3a_plantingw3 sectc3b_plantingw3 sectc3c_plantingw3
C4	Land Prices and Credit	sectc4a_plantingw3 sectc4b_plantingw3

Table 7.3b: Post-harvest Community datasets

Section	Section Name	Dataset Filename
Cover	Cover	sectc_harvestw3
C1	Respondents Characteristics	sectc1_harvestw3
C2	Community Infrastructure and Transportation	sectc2_harvestw3
C3	Community Organizations	sectc3_harvestw3
C4	Community Resource Management	sectc4_harvestw3
C5	Community Changes	sectc5_harvestw3
C6	Community Key Events	sectc6_harvestw3
C6A	Conflict	sectc6a1_harvestw3
		sectc6a2_harvestw3
		sectc6a3_harvestw3
C7	Community Needs, Actions, and Achievements	sectc7_harvestw3
C8	Food Prices	sectc8a_harvestw3
		Sectc8b_harvestw3

7.2.5 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. See Appendix 4 and the next section on the geo-variables which are made available in lieu of actual locations of household dwellings and plots.

7.2.6 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 4 on how these variables are constructed and linked to the GHS-Panel data. The table in Appendix 4

provides the name, type, source, reference period, resolution, description, and source of each geospatial variable included.

7.2.5 Status of household and individuals

Two additional data sets are released with Wave 3 which summarize the status of households and individuals across all six visits of the three waves: ***HHTrack.dta*** and ***PTrack.dta.*** The HHTrack data set also contains the full set of survey weights (see Section 3.0 for details on the weights).

7.2.7 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 sets of conversion factor files are included with the data. The first is ***food_conv_w3.dta*** which contains the conversion factors for food quantities in the food consumption file. The second is the dataset ***ag_conv_w3.dta*** which contains conversion factors for crops to be used with the agricultural module. For more information on these files and how to use them, see Section 8.4.1.

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 2. If questions were dropped in Wave 3, 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***. So 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 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. 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.

Location variables: ***zone***, ***state***, ***lga***, ***sector***, ***ea*** and ***ric*** have not been included in all the datasets. Instead, these variables have been included in the questionnaire cover datasets, i.e. secta_harvestw2, secta_plantingw2, sectc_harvestw2 etc., and from there they can be merged into any of the other datasets using the key variables as explained above.

8.3 Network Roster

A network roster is included in both the post-planting and post-harvest agriculture questionnaires. The network roster keeps a record of the list of places (businesses, markets, persons etc.) with which the household engages in agricultural trading activities. Each place is assigned the network code of the line in which it is in that section. Each place is recorded only once so we have for example, network codes N1, N2 etc. which is just a serialization of the places. This is similar to the household roster where an individual acquires the individual code of the line in which the person's name is written.

After the information has been entered in the network roster, the network code can be used in any section of the Agriculture Questionnaire where a place of trading is requested. The network roster contains information on the type of place and its location.

8.4 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 3, 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.

8.4.1 Unit Conversion Factors

The expanded list of units used in Wave 3 required additional conversion factors not previously available to convert these non-standard units into a common standard unit (kilograms or litres). In order to collect the item-unit weights required to calculate conversion factors, a specialized market survey was implemented prior to commencement of Wave 3. Reference photographs were also taken for all item-unit weights collected. The market survey was conducted in all 6 geopolitical Zones (2 States in each Zone) in an effort to capture variations in conversion factors throughout the country.

A wide array of item-unit weights was collected in this survey and were then used to calculate conversion factors. The calculated conversion factors are contained in ***food_conv_w3.dta*** and ***ag_conv_w3.dta*** included in the Wave 3 data. In both files, there are separate variables which have zone-specific conversion factors (e.g. *conv_NC_1*). There is also a national conversion factor (*conv_national*). Where conversion factors were acquired for a particular zone, the average conversion was included for the zone. However, if there was no conversion found in a zone, the national average was used for the zone-specific conversion variables. Although these conversion factors 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.

In order to use the conversion factors, one has to multiply a crop or food item with a conversion factor. For example, the dataset ***sect7b_plantingw3.dta*** features question 2, which asks how much the household consumed of each food item. One household is said to have consumed 1.5 large heaps of onions. In order to convert "large heap" to kg, the dataset ***food_conv_w3.dta*** has to be merged on the item code and unit code, and then the quantity (1.5 in this example) is multiplied with the relevant conversion factor. This could either be the conversion factor for that household's particular zone (variable *conv_SE_4* for South East) or the national conversion factor (variable *conv_national)*. It is highly recommended that the zone specific conversion factors are used. The same procedure can be followed to convert crop quantities using ***ag_conv_w3.dta.***

8.4.2 Reference Photo Album

Although reference photographs were used in Waves 1 and 2 of the GHS-Panel, in Wave 3 the collection of photos was greatly expanded and improved. The photos were collected a systematic manner during the market survey 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 (filename "X") 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 3

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 Tracking

One challenge was the way in which the interviewers completed the tracking forms. In some instances, the tracking form was not properly completed and this resulted in significant difficulty, and even failure, in tracking the relocated households. During the post-harvest training, a greater emphasis was placed on properly filling out these forms. As a result, filling of the tracking forms was improved in the post-harvest visit, making the tracking exercise easier to implement. In future Waves, the importance of filling tracking forms completely should be emphasized in the training.

9.2 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. Nonmeasurement 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 a fair distance 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: Land*) 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 92 percent. Some plots could still not be measured due to flooding or security concerns in some areas. In future waves, the rate of GPS measurement by each team must be more closely monitoring in the post-planting visit to ensure that plots are consistently being measured.

9.3 Availability of Electricity

Electricity was required by the data entry operator to operate the laptop computer and printer when in the field. This problem was anticipated so provision was made for use of generators in the State offices and while in the field. This strategy was partially successful, but there still some challenges encountered. In some cases, when teams were far from the State office it was difficult to locate a generator to use. As a result, the data entry operator had to return to the State office to recharge and continue entering questionnaires before re-joining the team in the field.

9.4 Security Problems

The most significant challenge faced during Wave 3 was the security situation in the North East of the country, particularly Borno and Yobe states. In both states, there were several areas that

were impassable due to road blocks by security forces or were deemed too hazardous for field staff to visit. A total of 14 EAs (12 in Borno State and 2 in Yobe State) could not be visited and thus no households were interviewed in these EAs. 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.

9.5 Data Entry

Data entry has been a significant source of errors in previous waves of the panel survey. Following the recommendation from Wave 2, in Wave 3 a double data entry system was implemented. All household and agriculture questionnaires were entered twice: once in the field and again at NBS headquarters. This design was meant to correct any data entry errors from the first data entry ensure that the data entered exactly matches what was written in the questionnaire and therefore. However, there were some challenges encountered in this process. The most significant challenge was that some interviewer errors identified while in the field were corrected in the data entry system, but not effected on the questionnaire. These errors were therefore still present when entered during SDE. This required a more deliberate review of both FDE and SDE data during the cleaning process (see the next section).

If a double data entry system is used in the future, it must be further emphasized to interviewers that all correction made in the data entry system should also be effected in the questionnaire. In the end, what is entered in the data entry system should correspond exactly with the questionnaire.

Appendix 1: How to Obtain Copies of the Data

The data are available through the NBS web site:

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

or through the LSMS-ISA website:

http://www.worldbank.org/lsms-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 Poverty and Inequality Division of the World Bank (see address below); and (c) not pass the data to any third parties for any reasons.

Leo Sanni
Statistical Information Officer
Plot 762, Independence Avenue,
Central Business District,
FCT, Abuja
Nigeria
www.nigerianstat.gov.ng
Phone: +2348033865388
Email: leosanni@nigerianstat.gov.ng

LSMS Database Manager
Poverty and Inequality Division
The World Bank
1818 H Street, NW
MSN MC3-306
Washington, DC 20433
www.worldbank.org/lsms-isa
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
1	0.00012	0.0027	0.00006
2	0.00016	0.004	0.00016
3	0.00011	0.00494	0.00004
4	0.00019	0.0023	0.00004
5	0.00021	0.0023	0.00013
6	0.00012	0.00001	0.00041

Note: All conversion is to Hectares

Appendix 3: Crop Codes

CROP	CODE	CROP	CODE	CROP	CODE
BEANS/COWPEA	1010	GINGER	2100	COCOA	3040
CASSAVA OLD	1020	GINGER PEELED	2101	COCOA POD	3041
COCOYAM	1040	GINGER SPLIT	2102	COCOA BEANS	3042
COTTON	1050	OTHER SPICES/VANILA	2103	COCONUT	3050
SEED COTTON	1051	GUM ARABIC	2110	COFFE	3060
COTTON LINT	1052	OKRO	2120	COFFE ARABICA	3061
COTTON SEED	1053	ONION	2130	COFFEE ROBUSTER	3062
GROUND NUT/PEANUTS	1060	PEPPER	2140	DATE PALM	3070
UNSHELLED GROUND NUTS	1061	SWEET PEPPER	2141	GRAPE FRUIT	3080
SHELLED GROUND NUTS	1062	SMALL PEPPER	2142	GUAVA	3090
GUINEA COURN/SORGHUM	1070	ATARE	2143	JUTE	3100
MAIZE	1080	PIGEON PEA	2150	KOLANUT	3110
UNSHELLED MAIZE(COB)	1081	PINEAPPLE	2160	KOLANUT UNSHELLED	3111
SHELLED MAIZE(GRAIN)	1082	PLANTAIN	2170	KOLANUT SHELLED	3112
POP CORN MAIZE	1083	POTATO	2180	BITTER KOLA	3113
MELON	1090	SWEET POTATO	2181	LEMON	3120
UNSHELLED MELON	1091	PUMPKIN	2190	LIME	3130
SHELLED MELON	1092	PUMPKIN LEAVE	2191	LOCUST BEAN	3140
WATER MELON	1093	PUMPKIN FRUIT	2192	MANDARIN/TANGERINE	3150
MILLET/MAIWA	1100	PUMPKIN SEED	2193	MANGO	3160
RICE	1110	GREEN VEGETABLE	2194	ORANGE	3170
UNSHELLED RICE(PADDY)	1111	DRY LEAVES(KUKA)	2195	OIL PALM TREE	3180
SHELLED RICE(MILLED)	1112	RIZGA	2200	FRESH FRUIT BUNCH	3181
YAM	1120	SHEA NUTS	2210	FRESH NUT	3182
WHITE YAM	1121	SOYA BEANS	2220	PALM OIL	3183
YELLOW YAM	1122	SUGAR CANE	2230	PALM KERNEL	3184
WATER YAM	1123	TEA	2240	AGBONO(ORO SEED)	3190
THREE LEAVE YAM	1124	TOBACCO	2250	OIL BEAN	3200
ACHA	2010	TOMATO	2260	PAWPAW	3210
BAMBARA NUT	2020	WALNUT	2270	PEAR	3220
BANANA	2030	WHEAT	2280	AVOCADO PEAR	3221
BEENI-SEED/SESAME	2040	ZOBO	2290	RUBBER	3230
CARROT	2050	ZOBO SEED	2291	RUBBER LUMP	3231
CUCUMBER	2060	APPLE	3010	RUBBER SHEET	3232
CABBAGE	2070	CASHEW	3020	CHERRY(AGBALUMO)	3240
LETUS	2071	CASHEW FRUIT	3021	ERU	3250
GARDEN EGG	2080	CASHEW NUT	3022	IYERE	3260
GARLIC	2090	CHILLI	3030		

Appendix 4: Confidential Information, Geospatial Variables

The GHS-Panel collects confidential information on respondents. The confidential variables pertain to (i) names of the respondents to the household and community questionnaires, (ii) village and constituency names, (iii) descriptions of household dwelling and agricultural plot locations, (iv) phone numbers of household members and their reference contacts, (v) GPS-based household and agricultural plot locations, (vi) names of the children of the head/spouse living elsewhere, (vii) names of the deceased household members, (viii) names of individuals listed in the network roster, and (ix) names of field staff. To maintain the confidentiality of our respondents, certain parts of the GHS-Panel database have not been made publicly available.

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 have been produced using the unmodified GPS data. Most of the underlying datasets are static (with exception of time-series), so the values should be largely unchanged relative to year 1, for non-mover households. Note that there may be some variation due to GPS data entry error, differences in data collection procedure, and technical limitations of the device. Geospatial variables are provided in 2 separate files: NGA_PlotGeovariables_Y2 and NGA_HouseholdGeovariables_Y2.

NGA_PlotGeovariables_Y2

The household plot-level file, *NGA_PlotGeovariables_Y2*, 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.

NGA_HouseholdGeovariables_Y2

The household-level file, *NGA_HouseholdGeovariables_Y2*, contains a range of variables measuring (on the basis of the household dwelling) 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 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 second 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 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 data are unchanged. The three distance variables have been updated using more reliable spatial datasets.

Table A4.1 NGA_PlotGeovariables_Y2

Theme	Source	Dataset Title	Variable Name	Variable Type	Reference Period	Resolution	Description	Web
Distance	LSMS-ISA	Plot Distance to Household	dist_household	Continuous	N/A	N/A	Plot distance to household	
Soil & Terrain	NASA	SRTM 90m	srtm_nga	Continuous	N/A	0.000833 dd	Elevation (m)	ftp://xftp.jrc.it/pub/srtmV4/arcasci/
	USGS	Slope (percent)	srtmslp_nga	Continuous	N/A	0.000833 dd	Derived from unprojected 90m SRTM using DEM Surface Tools	http://pubs.usgs.gov/of/2007/1188/, data provided USGS upon request
	AfSIS	Topographic Wetness Index	twi_nga	Continuous	N/A	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#

Table A4.2 NGA_HouseholdGeovariables_Y2

Theme	Source	Dataset Title	Variable Name	Variable Type	Reference Period	Resolution	Description	Web
Distance	FERMA	Household Distance to Main Road	dist_road2	Continuous	2013	N/A	Household 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_borderpost 2	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

Theme	Source	Dataset Title	Variable Name	Variable Type	Reference Period	Resolution	Description	Web
Landscape Typology	ESA and UC Louvain	GlobCover v 2.3	fsrad3_lcmaj	Categorical	2009	0.002778 dd	Majority landcover class within approximately 1km buffer	http://ionia1.esrin.esa.int/
	ESA and UC Louvain	GlobCover v 2.3	fsrad3_agpct	Continuous	2009	0.002778 dd	Percent under agriculture within approx 1 km buffer	http://ionia1.esrin.esa.int/
	IFPRI	IFPRI standardized AEZ based on elevation, climatology	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
Soil & Terrain	NASA	SRTM 90m	srtm_nga	Continuous		0.000833 dd	Elevation (m)	ftp://xftp.jrc.it/pub/srtmV4/arcasci/
	USGS	Slope (percent)	slopepct_nga	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_nga	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#
	LSMS-ISA	Terrain Roughness	srtm_nga_5_15	Categorical		0.000833 dd	Derived from 90m SRTM using 15 Meybeck relief classes and 5x5 pixel neighborhood	
	FAO	Harmonized World Soil Database	SQ1	Categorical		0.083333 dd	Nutrient availability	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	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/
	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	Rainfall Estimates (RFE)	anntot_avg	Continuous	2001-2012	0.1 dd	Average 12-month total rainfall (mm) for Jan-Dec	ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/
	NOAA CPC	Rainfall Estimates (RFE)	wetQ_avg	Continuous	2001-2012	0.1 dd	Average total rainfall in wettest quarter (mm) within 12-month periods from Jan-Dec	ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/
	NOAA CPC	Rainfall Estimates (RFE)	wetQ_avgstart	Continuous	2001-2012	0.1 dd	Average start of wettest quarter in dekads 1-36, where first dekad of Jan =1	ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/
	NOAA CPC	Rainfall Estimates (RFE)	h2012_tot	Continuous	2012	0.1 dd	12-month total rainfall (mm) in Jan-Dec, starting January 2012	ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/

Theme	Source	Dataset Title	Variable Name	Variable Type	Reference Period	Resolution	Description	Web
	NOAA CPC	Rainfall Estimates (RFE)	h2012_wetQ	Continuous	2012	0.1 dd	Total rainfall in wettest quarter (mm) within 12-month periods starting January 2012	ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/
	NOAA CPC	Rainfall Estimates (RFE)	h2012_wetQstart	Continuous	2012	0.1 dd	Start of wettest quarter in dekads 1-36, where first dekad of January 2012 =1	ftp://ftp.cpc.ncep.noaa.gov/fews/newalgo_est_dekad/
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	eviarea_avg	Continuous	2001-2012	0.004176 dd	Average total change in greenness (integral of daily EVI values) within growing season, averaged by state	ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005
		MOD12Q2 Land Cover Dynamics (PHENOLOGY)	evimax_avg	Continuous	2001-2012	0.004176 dd	Average EVI value at peak of greenness, averaged by state	ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	grn_avg	Continuous	2001-2012	0.004176 dd	Average timing of onset of greenness increase in day of year 1-356, within early growing season, averaged by state	ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	sen_avg	Continuous	2001-2012	0.004176 dd	Average timing of onset of greenness decrease in day of year 1-356, within growing season, averaged by state	ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	h2012_eviarea	Continuous	2012	0.004176 dd	Total change in greenness (integral of daily EVI values) within growing season of 2012, averaged by state	ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005

Theme	Source	Dataset Title	Variable Name	Variable Type	Reference Period	Resolution	Description	Web
		MOD12Q2 Land Cover Dynamics (PHENOLOGY)	h2012_evimax	Continuous	2012	0.004176 dd	EVI value at peak of greenness within growing season of 2012, averaged by state	ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	h2012_grn	Continuous	2012	0.004176 dd	Onset of greenness increase in day of year 1-356, within growing season of 2012, averaged by state	ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	h2012_sen	Continuous	2012	0.004176 dd	Onset of greenness decrease in day of year 1-356, within growing season of 2012, averaged by state	ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005

Appendix 5: Sampling Details

Final Weighting Procedures for Nigeria Panel Survey by Wave and Visit

1. Background

The National Bureau of Statistics (NBS) of Nigeria has been conducting the General Household Survey (GHS) annually in recent years to measure various socioeconomic indicators at the state level. The Nigeria Panel Survey is a longitudinal survey based on a subsample of the primary sampling units (PSUs) and households selected for the GHS. The integration of the longitudinal Panel Survey with the GHS makes it possible to conduct a more comprehensive analysis of poverty indicators and other socioeconomic characteristics, and improve the precision of estimates of trends in the indicators over time.

The Panel Survey is conducted annually as a longitudinal survey, with two visits each year, scheduled based on agricultural seasons in order to obtain information on the crop production during the second visit. When a sample household moves the survey includes procedures to tract the household in order to collect the data from that household at its new location. However, in the case of split households where only some of the household members move, no tracking is conducted.

The first (baseline) Panel Survey was conducted in the second half of 2010, referred to as Wave 1, Visit 1. For the current analysis of the Panel Survey data there are datasets available for each of two visits from the first three waves. Although some tables will be produced for the data from each individual visit, the data from different visits will also be matched to obtain files with the combined data for each wave, as well as a combined data file for all visits of the different waves. Since the response rate is different for each visit, the combined data files are limited to the households with completed interviews for all visits. For this reason, it is necessary to calculate different weights for each visit, and for the data from each combination of visits.

The purpose of this report is to describe the weighting methodology for the Nigeria Panel Survey, including all the procedures involved in calculating the weights for each wave and visit, as well as for the combined datasets. Since the weights are based on the sample design, the next section provides a brief description of the sample design for the GHS and the Panel Survey. The panel of sample households is established during the baseline survey, so the basic weights are determined based on the sampling frame for that survey. The weights for other visits and combined datasets are then adjusted based on the interview results for the individual visits.

This report was updated for Wave 3 of the Panel Survey to document the calculation of the weights for that round and for the matched panels from previous rounds. In Wave 3 a conflict module was added to the Panel Survey. The sample EAs in the Panel Survey that are located in areas classified as conflict areas were identified, and a supplemental sample of 5 households was selected in each of these EAs, after a new listing was conducted. The purpose of this expanded sample was to increase the number of sample households in the conflict areas in order to improve

the precision of indicators related to the effects of the conflicts. The weighting procedures for this full sample including the additional households in the conflict areas are also described in this report.

2. Summary of Sample Design for the 2010 GHS and Panel Survey

A comprehensive description of the sample design for the 2010 GHS and the Panel Survey can be found in the report on "Final Sample Design and Estimation Procedures for 2010 Nigeria General Household Panel Survey" (Megill, July 2010). A summary of the sample design is presented here in this section. Since the Panel Survey is based on a subsample of the 2010 GHS sample primary sampling units (PSUs) and households, the sample design for the GHS is described first.

A multi-stage stratified sample design was used for the GHS and the Panel Survey. These and other national household surveys conducted by the NBS are based on a master sample referred to as the National Integrated Survey of Households (NISH). The sampling frame for the NISH was based on the list of enumeration areas (EAs) defined for the 2006 Nigeria Census of Population. The EAs are operational segments defined for the census enumeration, with an average of about 34 households each. The EAs are identified on maps with well-defined boundaries. The EAs are used as the primary sampling units (PSUs) selected at the first sampling stage for the national household surveys. The institutional population living in prisons, hospitals, military barracks, school dormitories, etc. are excluded from the universe defined for the household surveys. The average number of households per EA is about 37 for the urban areas and 33 for the rural areas, and the variability in size is somewhat higher for the urban EAs.

The NBS has classified the EAs in the NISH master sample by urban and rural sectors, based on characteristics of the infrastructure. However, this information was not used for stratifying the master sample, since it is not considered official. The urban and rural sectors are used as domains of estimation for most national household surveys in Nigeria, although in the past some surveys have not always used a consistent definition of the urban and rural classification.

First the NBS selected a master sample of EAs in each Local Government Area (LGA) that could be used for any LGA-level survey. For this LGA master sample 30 EAs were selected with equal probability within each LGA for the 36 states, and 40 EAs were selected in each LGA for Abuja FCT. There are 769 LGAs in the 36 states of Nigeria and 6 LGAs in Abuja, so a total of 23,310 EAs were selected for the LGA master sample. The 30 EAs selected for the master sample in each LGA were divided into 10 systematic replicates, identified in the frame by the replicate identification code (RIC).

The individual states of Nigeria are the geographic domains of analysis for most of the national household surveys such as the GHS. Therefore, the NBS has selected the NISH sample EAs as a state-level subsample of the LGA master sample EAs, consisting of 200 sample EAs for each state. In order to select the NISH subsample of EAs in each state, the 30 master sample EAs in each LGA for that state were pooled together. The total number of EAs in the LGA master sample for each state is equal to 30 times the number of LGAs in the state. Then a systematic sample of 200 sample EAs was selected with equal probability at this second stage across all LGAs within

the state. The NISH sample EAs in each state were divided into 20 replicates of 10 EAs each. These replicates are identified in the NISH sampling frame by the NISHRIC. The sample EAs for most national household surveys such as the GHS are based on a subsample of the NISH master sample, selected as a combination of replicates from the NISH frame.

The GHS is based on a subsample of replicates from the NISH frame. A total of six NISH replicates with 60 EAs for each state was selected for the GHS. For GHS 2010 the sample EAs in the selected replicates for each state are identified in the frame with NISHRIC 10 to 15. The listing for the GHS sample EAs is supposed to be updated each year, but apparently this is not implemented consistently because of limited resources. At the next sampling stage 10 households are selected for the GHS in each sample EA systematically with equal probability from the listing. In this way a total of 2,220 EAs and 22,200 households were selected for the GHS.

The geographic domains of analysis for the Panel Survey are the six geo-political zones of Nigeria. Therefore it was necessary to have a sufficient sample size for each zone to ensure reliable estimates at this level. At the same time, the need for reducing the nonsampling errors and the plans for tracking individual sample households that move limited the sample size that can be managed with sufficient operational and quality control. Given these constraints, a maximum sample size of 5,000 households was considered for the Panel Survey.

The allocation of the 500 sample EAs and 5,000 sample households by zone for the Panel Survey is presented in Table A5.1, which also shows the distribution of the population by zone from the 2006 Nigeria Census results. Given that a similar level of precision is required for the Panel Survey estimates from each zone, a sample of 800 households was allocated to each zone except for the two largest zones, North-West and South-West, which were allocated a slightly larger sample of 900 households each.

Table A5.1: Distribution of population of Nigeria by zone, 2006 Nigeria Population Census, and allocation of sample EAs and households by zone for the Panel Survey

Zone	2006 Census population	%	No. of sample EAs	No. of sample households
North-Central Zone	20,369,956	14.5%	80	800
North-East Zone	18,984,299	13.5%	80	800
North-West Zone	35,915,467	25.6%	90	900
South-East Zone	16,395,555	11.7%	80	800
South-South Zone	21,044,081	15.0%	80	800
South-West Zone	27,722,432	19.7%	90	900
Nigeria	140,431,790	100.0%	500	5,000

Given the sample design for the GHS, the weights for this survey vary by the number of EAs in the LGA as well as the number of LGAs in the state. In order to stabilize these weights within each zone for the Panel Survey, the subsample of EAs for the Panel Survey were selected from the GHS sample EAs with probability proportional to size (PPS) within each zone, where the measure of size is based on the factor $N_{sh} * L_s$, where:

N_{sh} = number of EAs in 2006 Nigeria Census frame for LGA h of state s

L_s = number of LGAs in state s

As described later in the description of the weights, this sampling approach resulted in a sample that is approximately self-weighting within zone, as if the EAs had been selected with PPS directly at the first stage within each zone.

3. Reference Population for Longitudinal Survey

It is important to understand that the population of analysis for the Panel Survey corresponds to the sampling frame for the baseline survey, or Wave 1/Visit 1, in which the panel of sample households to be followed was determined. The panel does not represent newer households that came into existence after the baseline survey. In the case of split panel households, only the household in the sample EA with members who do not move is interviewed for the Panel Survey, including any new members of that household. When an entire sample household for the Panel Survey moved out of a sample EA, it was tracked. Therefore there is a one-to-one correspondence between the original panel households that moved and the tracked households. Since the Wave 1/Visit 1 weights were calculated and adjusted at the sample EA level, the weights for the other visits are also calculated at this level.

In the case of some longitudinal surveys such as the Tanzania Panel Survey, split households are also tracked. In this case it was necessary to apportion the original household weight based on the proportion of the household members that moved. Since split households were not tracked for the Nigeria Panel Survey, this simplified the weighting procedures for each visit following the baseline survey.

4. Calculation of Basic Weights for the GHS and Panel Survey

In order for the sample estimates from the data from each survey to be representative of the population, it is necessary to multiply the data by a sampling weight, or expansion factor. The basic weight for each sample household is equal to the inverse of its probability of selection (calculated by multiplying the probabilities at each sampling stage). Since the Panel Survey is based on a subsample of the GHS sample, the probability of selection of the households for the GHS is presented first. Based on the sample design for the NISH and the GHS, the probability of selection for the GHS sample households can be defined as follows:

$$p_{shi} = \frac{30}{N_{sh}} \times \frac{200}{30 \times L_s} \times \frac{60}{200} \times \frac{10}{M_{shi}} = \frac{600}{N_{sh} \times L_s \times M_{shi}}$$

where:

p_{shi} = probability of selection for the GHS sample households in the i-th sample EA in LGA h of state s

N_{sh} = number of EAs in 2006 Nigeria Census frame for LGA h of state s

L_s = number of LGAs in state s

M_{shi} = number of households listed in the i-th sample EA in LGA h of state s

The basic weight for the GHS sample households is the inverse of this probability of selection, calculated as follows:

$$W_{shi} = \frac{N_{sh} \times L_s \times M_{hi}}{600},$$

where:

W_{shi} = basic weight for the GHS sample households in the i-th sample EA in LGA h of state s

Since the Panel Survey is based on a subsample of the GHS, the probability of selection for the sample households in the Panel Survey has an additional component based on the subsampling rate. The sample EAs for the Panel Survey were selected with PPS, where the measure of size is calculated as $N_{sh} * L_s$. Therefore, the probability of selection for the Panel Survey sample households can be expressed as follows:

$$p_{zshi} = \frac{30}{N_{sh}} \times \frac{200}{30 \times L_s} \times \frac{60}{200} \times \frac{n_z \times N_{sh} \times L_s}{\sum_{s \in z} \sum_{h \in s} \sum_{i \in h} (N_{sh} \times L_s)} \times \frac{10}{M_{shi}} = \frac{600 \times n_z}{M_{shi} \times \sum_{s \in z} \sum_{h \in s} \sum_{i \in h} (N_{sh} \times L_s)} \quad ,$$

where:

p_{zshi} = probability of selection for the Panel Survey sample households in the i-th sample EA in LGA h of state s in zone z

n_z = number of sample EAs selected for the Panel Survey in zone z

The basic weight for the sample households selected for the Panel Survey is the inverse of this probability of selection, calculated as follows:

$$W_{zshi} = \frac{M_{shi} \times \sum_{s \in z} \sum_{h \in s} \sum_{i \in h} (N_{sh} \times L_s)}{600 \times n_z} \quad ,$$

where:

W_{zshi} = basic weight for the Panel Survey sample households in the i-th sample EA in LGA h of state s in zone z

It can be seen that the weights of the sample households for the Panel Survey in each zone vary only by number of households listed in the EA, confirming that the PPS subsampling procedures stabilize the Panel Survey weights within each zone.

Following the data collection for the baseline of the Panel Survey (Wave 1, Visit 1), the basic weights for this baseline visit were adjusted to take into account any non-interviews. The weights were adjusted for nonresponse at the sample EA level, as follows:

$$W'_{zi} = W_{zi} \times \frac{m_{zi}}{m'_{zi}},$$

where:

W'_{zi} = adjusted weight for the sample households in the i-th sample EA in zone z for the Panel Survey

m_{zi} = 10 = number of sample households selected in the i-th sample EA in zone z

m'_{zi} = number of sample households with completed Panel Survey interviews in the i-th sample EA in zone z for the Panel Survey

In order to simplify the terminology, the subscript *zi* will refer to the individual sample EAs selected within a particular zone (stratum); each sample EA is associated with a particular LGA and state within the zone.

In the case of the baseline Panel Survey (Wave 1, Visit 1), almost all of the non-interviews were replaced, so most of the non-interview adjustment factors were equal to 1. The weight adjustment factors for the subsequent visits were calculated in reference to the sample of completed interviews for Wave 1, Visit 1, which determined the panel of sample households to be followed, as described later in this report.

A spreadsheet for calculating the Panel Survey weights by sample EA was developed from the information in the sampling spreadsheet used for selecting the subsample of EAs from the GHS frame. Updated information on the number of households listed in each sample EA and the number of households with completed questionnaires for each visit were entered in this weighting spreadsheet, and formulas were used in the spreadsheet to calculate the weights.

5. Adjustment of Weights for Panel Survey Data Based on Population Projections

After the weighting spreadsheet was updated with the information for the replacement EAs and the geographic codes were consistent with those in the Panel Survey data file, the basic sampling weights were attached to household and person records in the data file for Wave 1, Visit 1. A tabulation of the weighted total population by state was produced from the survey data file to determine whether these results were consistent with the corresponding distribution from the 2006 Nigeria Census. It was found that there were considerable differences for some states, and the

weighted survey estimates were mostly lower than the census results, indicating that there was probably under-enumeration of households in the listing for many sample EAs. Conceptually, if there is a new accurate listing of households in each sample EA, the overall growth in the number of households across all the sample EAs would be reflected in the weighted estimates of the total population. The first set of design weights for the baseline Panel Survey were calculated in February 2011 and resulted in a weighted total population estimate of 120,282,653, based on using the original listing of households for the NISH. After that a new listing of households was conducted in the Panel Survey sample EAs in early 2012. The weights were re-calculated using the updated listing information, and the weighted estimate of the total population increased slightly to 122,242,860, while the 2006 Census population count was 140,431,790. Although the new listing improved the weights slightly, it appears that the new listing also undercounted the households in sample EAs.

One way to adjust the weights for such deficiencies in the listing is to use population projections at the state level for calculating weight adjustment factors. Estimates of the annual population growth rate by state were obtained from the National Population Commission (NPC) in order to calculate population projections by state. The information from the NPC included low, medium and high estimates of the population growth rate for each state; the medium growth rates were used for calculating the population projections. The reference date for the population projections is the mid-point of the data collection period for the 2010 Panel Survey for Wave 1, Visit 1. Since the survey was conducted between 1 September and 15 October 2010, the mid-point was determined as 23 September 2010. The calculation of the population projections by state involved using the population figures for each state from the 2006 Nigeria Census, and applying the annual population growth rate from the census reference date, 21 March 2006, to the survey reference date, assuming an exponential population growth rate. The following formula was used for calculating the population projection for each state:

$$\hat{P}_{S10h} = P_{C06h} \times e^{\ln\left[1+g_h \times \left(\frac{t_{S10}-t_{C06}}{365}\right)\right]}$$

where:

$\hat{P}_{S10h}$ = estimated total population for state h on 23 September 2010

P_{C06h} = total population for state h from 2006 Nigeria Census (21 March 2006)

g_h = annual population growth rate for state h

$t_{S10} - t_{C06}$ = number of days between the census reference date (21 March 2006) and the mid-point of the 2010 Panel Survey (23 September 2010), that is, 1647 days

The annual population growth rates were not available for the following states: Bayelsa, Ebonyi, Ekiti, Gombe, Nasarawa and Zamfara. Therefore, the national-level annual population growth rate of 2.86% was used for these states. Table A5.2 shows the 2006 Nigeria Census population

by state, the estimated population growth rates and the corresponding population projections by state for the mid-point of the data collection period for the baseline Panel Survey.

The weight adjustment factor based on the projected total population by state can be expressed as follows:

$$A_s = \frac{\hat{P}_{2010s}}{\sum_{i \in s} \sum_{j \in i} W'_{si} \times p_{sij}} ,$$

where:

A_s= adjustment factor for the weights of the baseline Panel Survey sample households in state s

W'_{si} = preliminary design weight for the sample households in the i-th sample EA in state s

p_{sij} = number of persons in the j-th sample household in the i-th sample EA in state s

The denominator of the adjustment factor A_s is the estimated total population in states from the baseline Panel Survey data using the preliminary weights. The preliminary weights for all the sample households within a state are multiplied by the corresponding adjustment factor for the state to obtain the final adjusted weight. After the adjustment factors are applied to the weights of each state, the weighted survey estimates of total population by state are consistent with the corresponding population projections.

Table A5.3shows the population projections and the weighted survey estimates of total population by state, and the corresponding weight adjustment factor for each state. The estimate of the total population of Nigeria based on the preliminary weights for the Panel Survey was 122,242,860, compared to the projected population of 160,243,147. It can be seen in Table A5.3that the weight adjustment factors vary from 0.6583 for Yobe to 2.9638 for Ekiti.

Table A5.2: Population distribution by state from 2006 Nigeria Census, estimated population growth rates and population projections for mid-point of baseline Panel Survey

State	Census Population 21-Mar-06	Annual Population Growth Rate (%)	Population Projection for Panel Survey 23-Sep-10
Abia	2,845,380	3.45	3,288,336
Adamawa	3,178,950	3.56	3,689,613
Akwa-Ibom	3,902,051	3.63	4,541,197
Anambra	4,177,828	3.06	4,754,691
Bauchi	4,653,066	3.27	5,339,641
Bayelsa	1,704,515	2.86	1,924,487
Benue	4,253,641	2.84	4,798,746
Borno	4,171,104	2.60	4,660,460
Cross River	2,892,988	3.36	3,331,606
Delta	4,112,445	3.29	4,722,961
Ebonyi	2,176,947	2.86	2,457,888
Edo	3,233,366	3.71	3,774,655
Ekiti	2,398,957	2.86	2,708,549
Enugu	3,267,837	3.52	3,786,881
Gombe	2,365,040	2.86	2,670,255
Imo	3,927,563	3.38	4,526,582
Jigawa	4,361,002	2.84	4,919,865
Kaduna	6,113,503	3.17	6,987,983
Kano	9,401,288	2.98	10,665,455
Katsina	5,801,584	3.15	6,626,212
Kebbi	3,256,541	3.10	3,712,073
Kogi	3,314,043	3.55	3,844,912
Kwara	2,365,353	3.39	2,727,176
Lagos	9,113,605	2.92	10,314,414
Nasarawa	1,869,377	2.86	2,110,625
Niger	3,954,772	3.31	4,545,449
Ogun	3,751,140	2.79	4,223,386
Ondo	3,460,877	3.05	3,937,184
Osun	3,416,959	3.28	3,922,684
Oyo	5,580,894	3.05	6,348,970
Plateau	3,206,531	3.13	3,659,408
Rivers	5,198,716	3.08	5,921,232
Sokoto	3,702,676	3.16	4,230,639
Taraba	2,294,800	3.22	2,628,228
Yobe	2,321,339	3.05	2,640,816
Zamfara	3,278,873	2.86	3,702,020
Abuja FCT	1,406,239	3.02	1,597,870
Nigeria	140,431,790	2.86	160,243,147

Table A5.3 Population projections and 2010 panel survey weighted estimates of total population by state, and corresponding weight adjustment factors

State	Population Projection 23-Sep-10	Preliminary Weighted Total Population	Weight Adjustment Factor
Abia	3,288,336	1,725,829	1.9054
Adamawa	3,689,613	4,800,971	0.7685
Akwa-Ibom	4,541,197	4,701,058	0.9660
Anambra	4,754,691	2,292,477	2.0740
Bauchi	5,339,641	5,806,461	0.9196
Bayelsa	1,924,487	845,021	2.2774
Benue	4,798,746	3,272,339	1.4665
Borno	4,660,460	4,895,182	0.9521
Cross River	3,331,606	2,770,295	1.2026
Delta	4,722,961	2,116,599	2.2314
Ebonyi	2,457,888	2,522,334	0.9744
Edo	3,774,655	2,105,676	1.7926
Ekiti	2,708,549	913,865	2.9638
Enugu	3,786,881	1,529,693	2.4756
Gombe	2,670,255	1,914,316	1.3949
Imo	4,526,582	2,317,123	1.9535
Jigawa	4,919,865	4,409,026	1.1159
Kaduna	6,987,983	5,684,699	1.2293
Kano	10,665,455	8,115,886	1.3141
Katsina	6,626,212	5,558,892	1.1920
Kebbi	3,712,073	4,235,476	0.8764
Kogi	3,844,912	1,630,927	2.3575
Kwara	2,727,176	3,490,907	0.7812
Lagos	10,314,414	4,738,437	2.1768
Nasarawa	2,110,625	1,994,670	1.0581
Niger	4,545,449	4,665,476	0.9743
Ogun	4,223,386	2,446,958	1.7260
Ondo	3,937,184	1,814,189	2.1702
Osun	3,922,684	2,563,124	1.5304
Oyo	6,348,970	3,735,603	1.6996
Plateau	3,659,408	4,084,778	0.8959
Rivers	5,921,232	3,752,908	1.5778
Sokoto	4,230,639	3,001,878	1.4093
Taraba	2,628,228	3,705,757	0.7092
Yobe	2,640,816	4,011,840	0.6583
Zamfara	3,702,020	2,550,275	1.4516
Abuja FCT	1,597,870	1,521,912	1.0499
Nigeria	160,243,147	122,242,860	

6. Nonresponse Adjustment of Weights for Panel Survey Data from each Visit and for Combined Datasets

The adjustment of the weights based on population projections established the final weights for the baseline Panel Survey (Wave 1, Visit 1). The total number of completed household interviews for the baseline survey at that time was 4,987. However, when the new set of weights was being calculated for each visit of the Panel Survey in September 2013, it was found that the Wave 1/Visit 1 data actually had 4,997 completed household interviews. Therefore, a very slight adjustment was made to correct the final baseline weights for a few EAs with a small difference in the number of completed interviews. These baseline (Wave 1/Visit 1) weights were then used as the basis for calculating the weights for the remaining visits as well as for the combined dataset for each wave and a combined dataset for all visits of both waves. For each visit it was necessary to adjust the weights for nonresponse based on the final number of sample EAs covered in each zone, and the number of completed household interviews in each sample EA for that visit. The reason for this approach is that the baseline survey represents the national frame of households that is being followed each visit. In the case where a panel sample household moves, it is tracked; if it is not found or cannot be interviewed, the baseline weight for the corresponding EA is adjusted to reflect this type of non-interview as well as the non-interviews for sample panel households in that EA that did not move.

In the case of the two visits for Wave 2, there were a few sample EAs with no completed household interviews. There were five sample EAs with zero households interviewed in Wave 2/Visit 1 and three such sample EAs in Wave 2/Visit 2. In this case it was necessary to have an additional adjustment of the corresponding weights at the zone (stratum) level to take into account the sample EAs without survey data. In the case of sample EAs that have zero households interviewed for a particular wave/visit, no weight appears in the weighting file. The weights will only be attached to the household records in each EA that have an interview status of 1 (completed) for the corresponding wave/visit.

In the case of Wave 2/Visit 1, there were 143 sample household records with interview status code 5, corresponding to "some sections during combined tracking"; this is the only wave/visit with such cases. The weights calculated for this visit and the combined files treated these sample households as non-interviews. If these partially completed interviews are included in the data for the analysis of particular sections of the questionnaire, conceptually it would be necessary to calculate a separate set of weights for each section. Since the poverty indicator and other consumption-related indicators depend on having complete consumption data, it would be necessary to exclude the households that are missing consumption data, so the weights should be calculated accordingly. For this reason, the final weights for the Panel Survey are only based on households with completed interviews.

The previous "population weights" in the Panel Survey data file were calculated in March 2012, based on 4,987 households with completed interviews for the baseline survey. As mentioned previously, there are 4,997 households with completed interviews in the current data file for the Panel Survey of Wave 1/Visit 1. The source of this difference is not clear, but this required a slight adjustment to the baseline weights so that the data for the 4,997 household records are

correctly weighted up to the frame level. These adjusted weights for Wave 1/Visit 1 became the base weights for the panel households. In order to calculate the weights for the subsequent visits, it was necessary to adjust the weights for each wave/visit for nonresponse at the EA level.

The final weights for each visit, calculated at the sample EA level, can be expressed as follows:

$$W_{(wv)zi} = W'_{(11)zi} \times \frac{m'_{(11)zi}}{m'_{(wv)zi}} \times \frac{n_z}{n'_{(wv)z}},$$

where:

$W_{(wv)zi}$ = final (adjusted) weight for the panel sample households in the i-th sample EA in zone (stratum) z for the Panel Survey in Wave *w* and Visit *v*

$m'_{(11)zi}$ = number of sample households with completed interviews for the i-th sample EA in zone z, for the baseline Panel Survey (Wave 1, Visit 1)

$m'_{(wv)zi}$ = number of sample households with completed interviews in the i-th . sample EA in zone z for the Panel Survey in Wave *w* and Visit *v*

n_z = original number of sample EAs selected for the Panel Survey in zone z

$n'_{(wv)z}$ = number of sample EAs in zone z with data (completed household interviews) for the Panel Survey in Wave *w* and Visit *v*

For the two visits of Wave 1 there were completed interviews in all the sample EAs, so the last adjustment factor was equal to 1 for all zones.

In the case of the Panel Survey data file for each wave with combined data from both visits, the analysis will be limited to the sample households that were successfully interviewed both times. In the same way, the weights for the combined data file from all waves and visits were based on the households with completed interviews for all visits. The weights for these combined files were also adjusted to represent the baseline frame.

In the case of the combined visits for each wave, the adjusted weight was calculated in a similar manner as follows:

$$W_{(w)zi} = W'_{(11)zi} \times \frac{m'_{(11)zi}}{m'_{(w)zi}} \times \frac{n_z}{n'_{(w)z}},$$

where:

$W_{(w)zi}$ = final (adjusted) weight for the panel sample households in the i-th sample EA in zone (stratum) z for the combined data from the Panel Survey for

both visits of Wave *w*

$m'_{(w)zi}$ = number of sample households with completed interviews for both visits of the Panel Survey for Wave *w*, in the i-th sample EA in zone z

$n_{(w)z}$ = number of sample EAs in zone z with data (completed household interviews) for both visits of the Panel Survey in Wave *w*

Finally, the weights for the combined data file of the Panel Survey for all waves and visits were calculated as follows:

$$W_{(all)zi} = W'_{(11)zi} \times \frac{m'_{(11)zi}}{m'_{(all)zi}} \times \frac{n_z}{n'_{(all)z}},$$

where:

$W_{(all)zi}$ = final (adjusted) weight for the panel sample households in the i-th sample EA in zone (stratum) z for the combined data from the Panel Survey for the visits from all waves

$m'_{(all)zi}$ = number of sample households with completed interviews for the visits from all waves of the Panel Survey, in the i-th sample EA in zone z

$n'_{(all)z}$ = number of sample EAs in zone z with data (completed household interviews) for the visits from all waves of the Panel Survey

These final weights for the combined Panel Survey datasets assume that the data are merged into one record for each sample household, so that the longitudinal analysis can examine trends in characteristics at the household level. However, if the data files are merged with separate household records for each visit, all the corresponding weights should be divided by the number of visits combined, so that total weighted number of households from the survey data will represent the frame for the baseline survey.

7. Calculation of Weights for Panel Survey Wave 3 and for the Combined Datasets

Since the same sample of panel households was maintained since the baseline Panel Survey, the panel weights for Wave 3 were calculated by adjusting the corresponding weights from Wave 1, Visit 1, based on the number of sample EAs that were covered in each zone and the number of households with completed interviews in each sample EA. This is the same methodology described above for the previous waves. The baseline panel weights had already been adjusted based on a new listing in 2012 as well as population projections. Given the longitudinal nature of the Panel Survey, the sample households still represent the frame at the time of the baseline survey.

Reference can be made to the formulas of the weights defined in the previous section. For each visit of Wave 3 it was necessary to compile data for the following components of the weight adjustment factors:

$m'_{(3v)zi}$ = number of sample households with completed interviews in the i-th . sample EA in zone z for the Panel Survey in Wave 3 and Visit *v*

$n'_{(3v)z}$ = number of sample EAs in zone z with data (completed household interviews) for the Panel Survey in Wave 3 and Visit *v*

The number of enumerated sample EAs for a particular visit of Wave 3 ($n'_{(3v)z}$) is obtained at the zone level, while the number of completed household interviews is counted at the sample EA level. This information was copied into additional columns of the previous weighting spreadsheet, followed by a column with the formula for the weight. Table A5.4 shows a summary of the total number of sample EAs and households with completed interviews for the combined visits of Wave 3. These results can be compared to Table A5.1 to examine the level of attrition in the panel.

Separate weights were also calculated for the combined data from different waves, for conducting the longitudinal analyses. In each case it was necessary to determine the number of sample EAs in each zone with interviews for all the combined waves, as well as the number of households in each sample EA that had completed interviews for all the combined waves. It should be noted that in Wave 3 an attempt was made to interview all the baseline sample households, even if some of them were not interviewed in Wave 2. The same formula was used for adjusting the baseline weights for nonresponse for each combined dataset.

Table A5.4 Distribution of the effective number of sample EAs and households with completed interviews by zone for the combined visits of Wave 3 of the Panel Survey

Zone	No. of sample EAs with panel data	No. of household interviews completed
North-Central Zone	80	777
North-East Zone	66	643
North-West Zone	90	882
South-East Zone	80	755
South-South Zone	80	744
South-West Zone	90	780
Nigeria	486	4,581

It can be seen in Table A5.4 that all of the sample EAs were enumerated in Wave 3 except for 14 sample EAs in the North-East Zone. A total of 4,581 panel households were interviewed in Wave 3 (both visits), indicating an attrition rate of about 8.4% since the baseline survey.