26 February, 2011

WORLD BANK HOUSEHOLD SURVEYS FOR THE AFRICAN MIGRATION PROJECT IN NIGERIA

<u>Methodology Report</u>

I.   Introduction

While migration is increasingly acknowledged as of great importance for development, migration data are rare and generally unreliable, leaving researchers and policymakers with weak bases for policy formulation. There are many anecdotal assessments of the direction, import and implications of both international and internal migration in developing countries, but little reliable data. Understandably, this has impeded the extent to which policy could harness and channel the immense resources involved in moving people and funds to achieve growth and development for the world's poorest. Clearly therefore, there is need for firmer grounds for understanding migration.

As part of a programme to improve data and understanding of migration in developing countries, the Development Prospects Group of the World Bank undertook in 2009 to conduct migration surveys in selected African countries, including Nigeria. The Nigeria project was implemented by Zibah Consults Ltd in partnership with resource persons from the University of Nigeria, Enugu Campus, and aimed to deliver representative data on migration for the entire country. Commissioned in September, 2009, it focused on understanding migration incidence (international and internal) as well as its role in development in Nigeria. Training for the enumerators and the fieldwork kicked off immediately. A change in methodology and the need to have a common methodology in the various countries participating in the World Bank project led to a recall of enumerators from the field for about 3 weeks after which they resumed fieldwork in October and concluded fieldwork in December 2009. Professor Richard Bilsborrow of the University of North Carolina and Mario Navarete, a sampling consultant with Sistemas Integrales, Santiago, Chile provided invaluable technical assistance to the Nigerian team, helping to draw up the methodological and sampling frameworks, supervising the output from the field and helping in organization of the dataset following the fieldwork. In addition, the entire team at the World Bank Development Prospects Group worked tirelessly to provide support needed to make the project a success.

Covering a country as large as Nigeria[1] with a sample size enough to allow for inference on migration incidence and its development impact demands a creative approach to sampling.

---

[1] With a population of about 150 million, Nigeria is the 9th largest country in the world and the largest country in Africa.

This is made the more difficult by the fact that in the present study, the project was limited to a sample size of about 2000 households. Consequently, a number of options were explored in designing the sample in order to obtain a nationally representative collection of households from which inference can be made on the incidence of migration, its determinants and its impacts, as well as flows and impacts of remittance on economic development.

This report aims to summarize the procedures, challenges and options used in the Nigeria migration survey to produce the dataset. It outlines the set of methodology and techniques employed, describe sample areas selected across the country and how they were sampled, and also describes the field team and fieldwork as well as problems with the methodology and difficulties in fieldwork. The aim of the report is to give the reader insight into the process that generated the data, the challenges faced and how they were resolved. It is hoped that, should the need for updating data arise in the future, any third party could pick up this report and use it as a guide to replicate the Nigerian project or even improve on it. This present report describes the project organization, sampling, fieldwork in the different zones of the country, and sample weighting and data quality issues.

II.   Project Organization and Pre-Fieldwork

A. Interviewer Training and Fieldwork Organization

After working to put together a highly qualified team for the project, the Project Management proceeded to organize a two-day training workshop to get the enumerators and supervisors to understand the purpose, style and outputs expected. The training further blended the team which had earlier been selected based on capacities expected to be critical to the successful collection of data from the target population. Some of the qualifications included being a graduate of an institution of higher learning (undergraduates with rich past experiences in field survey were considered in special cases) so as to have a solid education; being conversant in the local language or dialects of the area the enumerator applied to work in; and previous experience in similar surveys.

The training took place at the First Bank Learning Centre Enugu; a venue selected because it presents an excellent atmosphere with modern training facilities. Facilitators were selected based on their background and work experience. In addition to being accomplished researchers, they have all been previously involved in national surveys. Participation both by the resource persons and potential enumerators was quite intense and appreciable. All but one of the short-listed enumerators (who fell sick on the journey from Ilorin to Enugu) completed the training.

Training comprised a number of sessions. The first reviewed the background, objectives and rationale of the household survey of the African Migration Project, with the aim of getting the participants to appreciate the factors underpinning the project, existing data gaps, and implications for migration and remittance policies in developing countries. Thereafter, an overview of the proposed scientific methodology for the survey was presented and critically discussed, with very useful inputs from the participants leading to modifications and clarifications as deemed necessary. Then, for over a full day, the participants went through the survey instrument item by item, familiarizing themselves with the questions, amending questions to reflect the Nigerian context where appropriate, and discussing anecdotal responses and case studies. The project management structure, including reporting lines, roles and conduct of each team member, potential challenges and implications, documentation and field reports as well as relationships with respondents, was then discussed. Afterwards, those

that participated in the pilot study gave reports of their experiences and findings. These experiences and findings were deliberated upon, with participants weighing the implications of different challenges in the field and alternate means for dealing with them. The feedback also helped in further modifications of the instrument. Following this, the entire team of supervisors and enumerators were divided into groups for role playing. This further imprinted the lessons as different versions of potential problematic situations (some of which could not be dealt with by lectures and discussions) were discussed so enumerators could take steps to resolve them. This also helped highlight the time required to complete a questionnaire, different responses and attitudes of respondents, how to deal with them, among other things.

The project management team at Zibah Secretariat, headed by a project coordinator provided overall intellectual coordination for the project and took final decisions on all matters in the project, in consultation with the World Bank Migration team. The project coordinator was assisted by a project manager responsible for day to day management of logistics and associated issues. There was also a fieldwork coordinator who oversaw data gathering and later provided support for data cleaning and analysis. For each of the four zones in the sample (the North was merged into each zone with the change in methodology), a regional coordinator was appointed to oversee the work in the zone. The regional coordinator worked with the team leaders to ensure efficient and timely delivery of results from each sample state.

The fieldwork was planned to use an integrated system of interview and data entry with each team entering the information in the hard copy into soft copies at the end of each work day and submitting both hard and soft copies at the end of the fieldwork. Some teams were not able to enter the data at the end of each work day, while others did so at the end of the fieldwork. But these procedures did not yield optimal results as a substantial portion of the data had to be re-entered directly from the hard copy by data analysts at the end of the fieldwork before meaningful analysis could be done. There was continual feedback among all segments of the fieldwork teams and clarifications were sought and rendered on an ongoing basis. Changes in instructions quickly cascaded down the line and were promptly implemented. Daily communication from the secretariat through the supervisors to the enumerators ensured that this happened. In some cases though, hitches arising from poor telecom coverage hindered this communication; but overall, the process can easily be adjudged a great success.

B. Pilot Survey and Modifications to the Instrument

The pilot survey took place in two states – Enugu State in the Southeast region and Kogi State in the North Central region. The pilot survey was conducted in a total of 50 households from urban and rural areas of the two states. The survey was split equally between the two states with each taking 25. Interestingly, neither of the two states in the pilot study fell into the sample for the main study. The pilot study aimed to assess potential challenges of the survey; test the ease of usability of the survey instrument, and contribute to the training of enumerators. The project implementation team used expert knowledge to identify Enugu as a high migration incidence state and Kogi as a low migration incidence state. Two local governments – one rural and one urban were selected in each of the two states. For Enugu, Enugu North Local Government (an urban area) and Udi Local Government (rural area) were selected. One week was allocated for the pilot study but the actual fieldwork was from 25th through 28th August, 2009. The rest of the week was used by the teams to organize the responses and prepare a report on the experiences, including recommendations for the main survey. Each team in Enugu and Kogi consisted of two members. An overall coordinator

worked from Enugu to direct the teams and visit them in the course of the work. A training session was held on August 24 to prepare teams for the pilot study before they moved into the field.

No specific sampling framework was used to select sites (except that one had to be rural and the other urban) or households within the towns. Each team was allowed to randomly choose households within the towns given to them to cover. The Enugu team, for example, concentrated work on Abor town, Enugu metropolis, and Emene. Reports from the two teams showed that they got the cooperation of households only after a detailed explanation of the purpose and use of the survey.

Following the pilot, difficult segments of the questionnaire were identified and discussed. A number of questions in the questionnaire were also contextualized to suit expectations in the field. Areas where respondents had difficulties were identified, including age, household assets, finances and expenditures, among others. The pilot experience indicated there were some questions where respondents had disclosure difficulties. Dealing with such questions therefore featured prominently in the training. Disclosure difficulties were identified in questions involving age, household assets, finances and expenditures, among others. For example, in the rural areas, many household members either did not know their exact ages or were not eager to reveal them to strangers. Some gave ages that were doubtful given their looks, but of course, interviewers were not allowed to judge. They could only try to confirm from the respondents, but would not change the given ages even where they were doubtful. The same goes for household assets and finances which they considered personal and bordering on security. In some cases, responses given on personal finances and assets were thought not to reflect observable surroundings. In the discussions, interviewers were told to note exaggerated responses on assets and finances and compare them with surroundings. For example, households living in thatched huts but which claim to have fleet of cars are not to be taken very serious. Under such circumstances, the interviewer is expected to use his judgement to guide the respondent to more meaningful disclosures or ask counter questions in later stages of the interview which could lead to corrections of earlier, doubtful responses. While the expenditure and finance questions were considered standardized and easy to understand, interviewers were asked to introduce some of the questions as part of discussions on other issues. On the whole, participants were encouraged to use discussion approach and try to incorporate some of the more difficult questions as part of larger discussions. Under such circumstances, questions can even be skipped at some point where they are considered to be immature and brought in later as part of discussion or gleaned from responses to other questions.

The pilot experiences also had ample examples of hostile responses and poor cooperation from some respondents. A few cases were reported where a household head would neither cooperate with the teams nor allow members of his/her household to do so. The Enugu team reported one outstanding case in Abor where the household head (a woman) would not as much as allow the team near her compound. Reasons cited by persons in this class ranged from security to frustration with government of the day. It was observed in some cases that respondents mistook the teams for public officials on revenue drive or taking information that might be used someday for tax or other official purposes. They therefore would not want to give out information and sometimes would become outright hostile to the teams. Dealing with hostile respondents was therefore intensely treated in the training with many conceivable forms of hostility outlined and discussed. Innovative approaches and ideas for overcoming these challenges were examined. In particular, the idea of prior awareness was raised and

extensively discussed. It was agreed that wherever possible, teams should not begin interviews without meeting with the traditional institution (the Traditional Ruler and his Cabinet). Providing a written overview of the survey which goes with the instrument instead of having to rely on oral explanation from enumerators was also considered and adopted by the teams. The training sessions therefore emphasized the need for care, patience, clarifications to all possible bias of the respondent as means of winning his/her confidence.

The pilot study fed into the training. A session was provided in the final training for leaders of the pilot teams to provide feedback, and their recommendations were debated by the larger group. This feedback led to modifications of wordings or structures of the instrument, where considered appropriate to reflect the needs of the fieldwork. Also, indicative figures of the incidence of migration were obtained from the pilot, which helped in the design of the sampling framework for the main study.

III.   Sample Design

A. Sampling Frame

The sampling frame was the 2006 National Population Census. For administrative purposes, Nigeria has 36 states and the Federal Capital Territory. These states are grouped into six geopolitical zones – the North Central, North East, North West, South East, South South and South West. The states in turn are divided into 776 Local Governments. The demographic and political characteristics of the states vary considerably. For example, the number of component local government areas in the states ranges from 8 in Bayelsa State (in the South South) to 44 in Kano State (in the North West). Likewise state populations vary widely from 1.41 million in the Abuja Federal Capital Territory to 9.38 million in Kano State. The National Bureau of Statistics splits the country further into 23, 070 enumeration areas (EAs). While the enumeration areas are equally distributed across the local government areas, with each local government area having 30 enumeration areas, the differences in the number of local government areas across states implies that there are also huge differences in the number of enumeration areas across states. Appendix table 1 summarizes the population according to the 2006 population census (in absolute and proportionate numbers), number of local government areas, and number of enumeration areas in each state .

Given the above, a stratified random sampling technique was thought to be needed to select areas according to population and the expected prevalence of migrants. The National Bureau of Statistics (NBS) provided a randomly selected set of enumeration areas and households spread across all states in the Federation from the 2006 sampling frame. Every state in Nigeria has three senatorial zones (often referred to as North, Central and South or East, Central and West). The NBS sample enumeration areas were distributed such that within each state, local government areas from each senatorial zones were included in the sample, with Local Governments in each state nearly evenly distributed between rural and urban areas. In all, a total of 3188 enumeration areas were selected. These enumeration areas were unevenly spread across States; some states in the North West (Kano, Katsina, and Jigawa), and a few in the South South (Akwa Ibom and Delta) had over 100 enumeration areas selected while others such as Imo and Abia in the South East, and Borno, Gombe and Taraba in the North East, had as few as 20 enumeration areas selected. This selection partially reflected the relative population distribution and number of Local Government Areas in the component states. Annex Table B shows details of the states and geopolitical regions, their shares in population of the country, the number of Local Government Areas and enumeration

areas in each state and the number of enumeration areas given in the NBS list that formed the frame for the study.

B. The Sample for the Migration Survey

a. Sample Selection of States, Local Governments and Enumeration Areas

Originally, the intention was to have proportionate allocation across all states, using the population of each state in the 2006 Census to select the number of households to be included in the sample. But it was later recognized that this would not yield enough migrant households, particularly those with international migrants, especially as the total number of households that could likely be covered in the sample to was limited to 2000. Consequently, a disproportionate sampling approach was adopted, with the aim of oversampling areas of the country with more migrants. According to Bilsborrow (2006), this approach becomes necessary because migrants are rare populations for which a distinct disproportionate sampling procedure is needed to ensure they are adequately captured. Given the relative rareness of households with out-migrants to international destinations within the 10 year reference period (selected by the World Bank for all countries) prior to the planned survey, sampling methods appropriate for sampling rare elements were desirable, specifically, stratified sampling with two-phase sampling at the last stage.

Establishing the strata would require that there be previous work, say from the most recent Census, to determine migration incidence among the states. However, the needed census data could not be obtained from either the National Bureau of Statistics or the National Population Commission. Therefore, the stratification procedure had to rely on available literature, particularly Hernandez-Coss and Bun (2007), Agu (2009) and a few other recent, smaller studies on migration and remittances in Nigeria. Information from this literature was supplemented by expert judgement about migration from team members who had worked on economic surveys in Nigeria in the past. Information from the literature and the expert assessment indicated that migration from households is considerably higher in the South than in the North. Following this understanding, the states were formed into two strata– those with high and those with low incidence of migration. In all, 18 States (16 in the South and 2 in the North) were put into the high migration incidence stratum while 19 states (18 in the North and 1 in the South) were classified l into the low migration incidence stratum (column C of Appendix Table 1).

The Aggregate population of the 18 states in the high migration incidence stratum was 67.04 million, spread across 10,850 Enumeration areas. Thus, the mean population of an EA in the high migration stratum was 6179. In turn, the aggregate population of the 19 states in the low migration incidence stratum was 72.95 million spread across 12,110 EAs yielding a mean EA population of 6024. These numbers were close enough to assume the mean population of EAs was essentially the same. To oversample states in the high stratum, it was decided to select twice as high a proportion of the states as in the low stratum. To further concentrate the sample and make field work more efficient in being oriented to EAs more likely to have international migrants, we decided to select randomly twice as many LGAs in each state in the high stratum states as in the low stratum states.

Thus, 12 states were randomly selected with probabilities of selection proportionate to the population size of each state (so states with larger populations were accordingly more likely to fall in the sample) from the high stratum states. Then two LGAs were randomly selected from each sample state and 2 EAs per sample LGA (one urban, one rural) to yield a total of

12 x 2 x 2 or 48 EAs in the high stratum states. For the low stratum, 6 states were randomly selected. From each of these, 1 LGA was randomly picked and 2 EAs were selected per sample LGA to give a total of 6 x 1 x 2 or 12 EAs in the low stratum. This yielded a total of 60 EAs for both strata. Given the expected range of 2000 households to be sampled, approximately 67 households were to be sampled from each local government area or 34 households from each enumeration area.

So far, the discussion has assumed two groups of households – migrant and non-migrant households. However, the study was interested in not just lumping all migrants together, but rather in classifying migrants according to whether their destination was within or outside the country. Migrant households were thus subdivided into those with former household members who were international migrants and those with former household members who were internal migrants. Three strata of households were therefore required, namely:

1.   **Households with an international migrant:** at least one person who was a member of the household since Jan. 1, 2000 left to live in an international destination and has remained abroad;
2.   **Households with an internal migrant:** at least one person who was a member of the household since Jan. 1, 2000 left to live elsewhere in Nigeria (outside the sample LGA) and has not returned to the LGA; and
3.   **Households with no migrant:** No member of the household has left to live elsewhere either within or outside the country since Jan. 1, 2000.

The selection of states to be included in the sample from both strata was based on Probabilities of Selection Proportional to (Estimated) Size or PPES. The population in each stratum was cumulated and systematic sampling was performed, with an interval of 12.16 million for the low stratum (72.95 million divided by 6 States), and 5.59 million for the high stratum (67.04 million divided by 12 States). This yields approximately double the rate of sampling in the high migration stratum, as earlier explained. Using a random start between 0 and 12.16, the following states were sampled in the low stratum: Niger, Bauchi, Yobe, Kano, Katsina, and Zamfara. In the high stratum, states sampled were Abia, Ebonyi, Imo, Akwa Ibom, Delta, Edo, Rivers, Lagos, Ondo, Osun and Oyo[2]. Given its large population size, Lagos fell into the sample twice. The final sample, with LGAs and EAs moving from North to South (i.e. from the low to the high stratum states) is presented in Table 1 below.

The sample was concentrated in the South since that is where it was expected that more households have international migrants. It was expected that the survey would still also be reasonably representative of the whole country and of both internal migrant and non-migrant households through weighting the data. To this effect, field teams were asked to keep careful track at all stages of the numbers of people and households listed compared to the number in the actual sample in each stratum, at all stages of sampling; from the first stage of sampling states (the Primary Sampling Units, or PSUs), to Local Governments (LGs), and finally Enumeration Areas or EAs (see below). It is worth noting that the number of EAs to be selected from each sample LG is miniscule in every state compared to the total number of EAs in the state. Overall, the intended sample of about 2000 households would yield about 13 thousand persons, or only 0.00008 of the huge population of Nigeria. Eventually, a total of 2,251 households with 13,415 individuals were actually sampled.

---

[2] It is merely a coincidence that neither of the two high stratum states in the North nor the one low stratum state in the South was selected in the sample.

The next steps, then, were to select the local government areas (LGs) and enumeration areas (the last stage or ultimate sampling units, or UAUs), and then finally select households from the sample EAs in the selected states in each stratum. To select local government areas, it was decided a priori to maintain a balance between rural and urban areas. In the Northern States where there was to be only one LG selected per sample state, this selection was made randomly. Fortunately, with the urbanization rate at approximately 50 percent, there is a near balanced distribution of rural and urban areas in the country. The distribution between urban and rural areas of the selected local government areas was therefore fairly equal, as expected. For the Southern States, initially only two Local Governments were provided for in the sample, so an attempt was made to select one rural and one urban centre in each state. In some cases though, the distinction between urban and rural areas was blurred by insufficient information. It was also considered useful to ensure that the two local government areas selected fall in at least two different senatorial zones to improve representation of cultural and other differences among peoples in each state. Where such differences were considered not too significant to warrant special attention, equal representation of rural and urban areas was prioritized. Table 1 shows the States, Local Government areas and Enumeration Areas selected for the work.

b. Table 1: States, Local Government and Enumeration Areas in the Sample

A	B	C	D	E	F***
State	**Zone**	**LGs**	**Selected LGs**	**EAs**	**EAs selected**
Bauchi	NE	19	Ningi	570	2
Kano	NW	45	Minjibir Gaya**	1350	4
Katsina	NW	34	Kaita	1020	2
Niger	NC	25	Bida Chanchaga**	750	4
Yobe	NE	17	Gulani Gujba**	510	4
Zamfara	NW	15	Bungudu	450	2
Abia	SE	18	Umuahia North Bende Ikwuano**	540	6
Imo	SE	27	Owerri Municipal Ezinihite Owerri North**	810	6
Ebonyi	SE	12	Afikpo North* Ohaozara Ezza North**	360	6
Akwa Ibom	SS	31	Etinan Uyo Onna**	930	6
Delta	SS	25	Isoko South Warri S/West Isoko North**	750	6
Edo	SS	15	Esan North Owan East Ikpoba Okha**	450	6

Rivers	SS	23	PortHarcourt Oyigbo Obio/Akpor**	690	6
Lagos	SW	20	Ajeromi/Ifelodun Mushin Ikeja Badagry Oshodi/Isolo** Surulere**	600	12
Ondo	SW	19	Akoko S/East Idanre Akure South**	570	6
Osun	SW	30	Ife East Boluwaduro Ifelodun**	900	6
Oyo	SW	32	Ibadan North Ibadan S/West Ibadan N/East**	960	6
Total		407		12210	90

- *Selected communities in originally selected LG, Ishielu had communal crisis and could not be accessed by enumerators. A different LG had to be sampled keeping in mind the need to pick a local government within the same senatorial zone. Afikpo was therefore selected to replace Ishielu.
- **These are the additional Local government selected due to the increase in sample LGs by 50 percent.
- ***Includes the EAs in the additional LGs of column D.
- Note LGs and LGAs are used interchangeably in this document.

c. **Sample Selection of Households**

The mean population size of the EAs in Nigeria in the 2006 census was around 6000 persons in both strata (and assumed to be slightly larger by 2009). The number of households is therefore near 1000 on average, and always more than 100. Given available resources, it had been determined that it would be possible (and sufficient) to list only 100 to 150 households in each sample EA. To do this, each sample EA therefore had to be partitioned using a defined procedure into an average of 6 to 10 segments before, one of which is randomly selected. This ordinarily would require local maps or landmarks, perhaps a listing of dwellings, consultations with local government officials or police, etc. But there were no adequate maps for most areas. In fact, only major cities such as Lagos and Abuja had such maps, mainly street maps. Here, the NBS sample of 3188 enumeration areas was useful since maps were not available from previous surveys. The NBS sample contained sample listings of about 10 households in each of the 3188 enumeration areas in its national sample frame, obtained with the the intention that in a proportionate random sample across all states all ten would be interviewed. But in the present context, with the adoption of disproportionate sampling, the team found the randomly generated list of households useful only in locating a segment of each enumeration area. Thus for each sample enumeration area, about 90 adjoining households (the number depending on the team and the distribution of settlements in each enumeration area) were added to the list of 10 NBS selected households to make a total of 100 or more households to be listed. This made it unnecessary to develop complex partitioning directions for field teams. The 10 original households in the NBS list for each enumeration area might or might not be in the final sample given that they are only 10 of 100 or so listed households, and only a maximum of 34 could ever be sampled (see below). The teams were allowed to take on the nearest adjoining 90 households to the 10 selected. The count could go in any direction for highly populated places. In some cases where the

population of an EA is small, teams could take on other households that may not be exactly adjoining to make a list of 100 or so households. This way, it became irrelevant what the partitioning system could be for all households within the enumeration area and no new randomization or selection procedure for the partitioned enumeration area was needed.

To simplify weighting, it was decided to evenly allocate the sample of 2000 or so households across local government areas in the sample. For states with 2 local government areas, that meant having twice as much households in the sample compared to the states with only one local government area. Since most of the states with two or more local government areas are in the South with potentials for higher migration incidence, this achieves the basic purpose of oversampling households with migrants (international and internal). Dividing the total expected number of 2000 questionnaires by the 30 local government areas in the list gave an average of 68 questionnaires per local government area, i.e., 34 per EA. This was the original sampling plan, called Procedure A.

d. Household Listing and Sampling Procedures

Actual sampling of the households in the last stage of the 4-stage sample involved 2-phase sampling that, in the first phase, lists all households in a randomly selected part of the EA with about 100 occupied households (in both urban and rural EAs). Once a 'partition' was selected, the team listed the households in the partition, to show how many of the households were in the three strata – households with no (out-)migrant since year 2000, households with one or more internal migrants, and households with one or more international migrants. Households with no one in the eligible age group of interest, viz., thought to be involved in making their own migration decisions, taken to be persons aged 15 to 59, were excluded from the list of eligible households to be sampled. Thus only households with someone who had actually been aged 15 to 59 at the time of migration and who had left since January 1, 2000, and were still living away at the time of interview in 2009, were classified as households with migrants for the study. Households with only children under age 15 or persons aged 60 or more were recorded in the supervisor listing sheet summary but not eligible to be sampled even as non-migrant households. The reason is that the study of the determinants and consequences of migration—part of the goal of the survey was to provide data for that—would require that non-migrant households contain some adult aged 15-59, for comparison with the households which had someone leave in that adult age group. It was thus necessary to list more than 100 households, say 105 or 110 households, to have 100 at risk of having a relevant out-migrant or non-migrant. Even though such households are rare in Nigeria, this procedure was adopted to ensure an adequate sample size. The procedure also makes sense to allow for non-response.

Listing sheets and supervisor field control sheets were provided to team leaders. In addition, "supervisor sampling sheets" were prepared and provided to team leaders and supervisors for selecting (sampling) households from the three strata of households designated above.

Early in the fieldwork, it was determined that Procedure A based on the selection of 34 households from a listing of 100 to 150 per EA would not yield sufficient households with international migrants, and would, moreover, involve listing more households than necessary and hence involve longer and more costly fieldwork. This led to replacing Procedure A with Procedures B and ultimately C in the final stage of selecting households for interview in Enumeration Areas, the Ultimate Area Units (UAUs). Sections IVB,C describe and compare the three procedures below.

The number of households to be interviewed in a sample EA was originally set at 34 in Procedure A, with a fixed maximum of 12 households allocated to each of the two groups of households with international or internal migrants, and 10 to the households without migrants. Where the number of international or internal migrants was up to 12 in each group, all were selected automatically for interview. If the number was more than that, then 12 of each type of households were randomly selected from whatever the aggregate number was in the listing for each group. It was ensured that this was done while the team was in the EA, usually by the team leader, using a table of random numbers or other convenient random selection process. The balance, to achieve a total of 34 in the three strata combined, would be selected from the non-migrant household stratum. It was expected that there would always be sufficient number of non-migrant households. So, in an extreme case in which the quotas of 12 households each with recent international and internal migrants (as defined) were both filled, then only 10 households having no migrants need be selected for interview, to reach the maximum of 34 per EA. In such situations, the procedure would result in oversampling households with international migrants compared to those with internal migrants and non-migrants. But there was no a priori reason to expect that situations in which there were at least 12 households with recent international migrants would dominate. Thus if the listing produced, say, 2 households with international migrants, 40 with internal migrants, and 108 with no migrants, Procedure A would lead to samples from the three strata, respectively, of 2, 12 and 20, and hence few international migrants despite the large listing efforts.

With some concerns about non-response, it was allowed for teams to select up to 14 each from strata 1 and 2, and 11 from stratum 3, to ensure getting at least 12, 12, and 10 completed responses in most situations. Shortfalls in stratum 1 and 2 would be made up by increasing the number interviewed in stratum 3. This means a final range of 0 to 12 households in stratum 1, 0 to 12 households in stratum 2, and 10 to 20 households in stratum 3, in each EA (the maximum in stratum 3 being fixed at 20). The result of this process would be to select and hopefully interview up to 34 households in each sample EA. The total number of households in the survey would then be about 2000 or more. However, a major problem with this approach is that it guaranteed sampling and interviewing more internal migrants than international migrants, since there would be far more cases in which there would not be 12 households listed with qualified international migrants than 12 households with internal migrants.

IV. Summary of Fieldwork

A. Fieldwork in the Different Zones – Matters Arising

The fieldwork was originally scheduled to be completed between September 7 and September 30, 2009. It actually began on 14 September but significant changes in the methodology warranted that the teams be withdrawn from the field while the sample and implementation strategy was re-drawn (see above, the evolution from Procedure A to Procedure B to Procedure C). Following this and the adoption of the new sampling framework, the teams headed back to the field on the 5th of October, with a mandate to finish within three weeks, but the difficulties associated with getting to some of the sample communities were greater than anticipated, leading most of the teams to spend far more time in the field than was projected.

For states in the South East - Imo, Ebonyi and Abia states – led by Chioma Onwumelu, the interviews were ultimately conducted on a total of 405 households: 106 non-migrant

households, 160 households with internal migrants, and 139 households with international migrants. (Details are found in Annex A.) These households were selected from a sampling frame near evenly distributed between the population living in urban and rural areas. The number sampled represents approximately 24 percent of the *1679 households listed by teams in the* selected sample *states of the zone.* Almost all targeted households responded. A few cases required the intervention of the supervisor with the help of community authorities to encourage sample households to respond. Given the different teams' understanding of the local language and socio-cultural terrain of each region, only in rare cases were there challenges in communication between the team and respondents. Cooperation and understanding within the team was appreciably high. For example, when the team in Imo finished earlier than those in Abia and Ebonyi, they agreed to be posted to those states to assist with the work at no extra cost to the project. This was also due to the sterling leadership of the regional coordinator who had meetings with the teams before they departed to the field. During the meetings, each field worker was brought to see the success of the region as the success of his own state and therefore the success of his/her own fieldwork, creating both collegiality and collective responsibility. This proved invaluable at the different stages of changes in the field methodology described above, when exchanges of information among teams enhanced overall performance of teams in the region.

Despite this, the team confronted a number of challenges. For example, in some instances there was difficulty in obtaining correct information in sections of the questionnaire seeking personal (particularly financial) data. Understandably[3], many households were apprehensive about divulging information relating to household finances. In some cases, letters of introduction and identity cards issued to the enumerators and supervisors were not sufficient to convince these households; it took lengthy pleas and oaths from the fieldworkers to convince them to respond, even though most teams had already met with traditional local community heads and paid for publicity using town criers. In a few others when practicable, inferences were made from information on daily expenditures and what could be gleaned from the residence and its contents to estimate weekly or monthly expenditures.

Accessibility of some selected communities was also an issue, particularly in Ebonyi state where infrastructure is weaker than in the other states of the region, making it difficult to access some enumeration areas. Ezillo community in Ishielu Local Government (of Ebonyi State) was originally selected but involved in tribal wars so had to be dropped and replaced by another community with Oriuzo in Ezza North from the sample in order to not jeopardize the security of interviewers. Some areas were not easily accessible by car so interviewers had to use motor bikes or trek very long distances.

In the four states of the South South – Akwa Ibom, Edo, Delta and Rivers – led by Eric Onyebalu, a total of 1,642 households were listed out of which 523 or 32% were interviewed (see details for each state in Annex A). The teams were directed by the supervisor to first conduct physical identification of the enumeration areas. Thereafter, they had comprehensive listings of the households in the first set of enumeration areas (this was before the 50 percent increase in Local Government and Enumeration area coverage). Like the team in the South East, the South South team also demonstrated high mutual respect for one another which made it possible to overcome the many challenges they faced. Adherence to the provisions of

---

[3] Security risks across the country are not negligible and therefore household often are apprehensive divulging personal information to strangers. This is not peculiar to the migration survey. Most all surveys in the country report same issues.

the codes of conduct and mutual desire to resolve differences in team members' perceptions of the work proved critical in keeping them together and yielding regular progress.

Given that the communities and enumeration areas selected for the project in the states were randomly chosen, they were far apart, leading to significant logistic problems. In a number of cases, sample communities were not covered by mobile telephones, raising difficulty in communication between teams and the supervisor(s), on the one hand, and among the teams, on the other. Some of the states in the South South (the Niger Delta area) are volatile because of activities of militants. So security was a major concern in this zone. There were also challenges with transportation emanating from swampy neighbourhoods. At one point, a sample EA (and consequently the Local Government, Uvwie) had to be replaced with another one (Isoko South) on account of the combined challenges of security and navigability. Fortunately, the survey took place in the dry season, reducing logistic difficulties so interviews were successfully held in the rest of the LGs and EAs.

As in the South East, many respondents were not comfortable with some questions on personal incomes and family finances. The teams therefore had to use tact to extract relevant responses. To this end, examples cited and instances from role playing helped significantly to deal with such challenges in the field. Meanwhile, following the understanding that international migrants should be oversampled, the team most often employed their knowledge of the local environment and the reconnaissance listings to explore means of reaching households with international migrants. As all teams were required to deliver soft as well as hard copies of the responses, the South South team spent a lot of time following the end of fieldwork to translate responses on papers into soft copies. The reason for this is that it was not possible for them to key in data from the questionnaires at night due to the rigorous work of interviews. After observing the difficult work for the first few days, the regional supervisor excused the teams from the task of daily keying in responses. Fortunately, the team was also as diligent after the fieldwork to use the first couple of days in filling out the soft copies of the instruments. Overall, the fieldwork in the region was a success.

Sample states in the North included Kano, Katsina and Zamfara (in the Northwest), Bauchi and Yobe (in the Northeast) and Niger (in the North Central) geopolitical zones. Given the much lower out-migration rate in this part of the country, analyses on this study often treat them as a homogenous group. Fieldwork in the three regions was coordinated by Uchenna Amaeze. Much more than other parts of the country, the Northern countryside is characterized by states with large land masses and great distances between towns and communities. Some are so far apart that moving from one sample EA to the next could mean several hours' drive, requiring a relocation of the entire team and materials. Invariably, the result was a lag in work that could be up to three days before teams could settle down well enough to resume listing and interviewing in a new EA. As in other parts of rural Nigeria, some communities in the sample were not accessible by motor vehicle, so teams had to severally use motorbikes in areas not accessible to cars.

Teams in the North also experienced additional challenges with the relatively lower rate of literacy in sample areas. In some EAs, the use of local consultants notwithstanding, it took an unusually long time to finish a single interview as the interviewers had to spend a lot of time explaining to respondents the information they were asked to supply and why. This resulted in longer interview completion times, return visits to households for reconfirmations. Otherwise, respondents were generally very co-operative, not unconnected to the fact that most interviewers working in the region were from the region and had considerable

experience in conducting surveys there. The teams also made things easier by paying prior courtesy calls to the local authorities and chiefs, who then introduced them to the community, soliciting that the people give them maximum co-operation.

The teams in the North also showed tremendous understanding given the myriad challenges that cropped up during the exercise. Five members of the team (independently and at different times and in different states) were involved in serious road accidents as they travelled through their designated EAs; one required medical attention and had to be replaced, others took a few days off, after which they were able to return to their work. Being far from project headquarters and far from one another in remote and hard-to-reach communities made it difficult for the regional supervisor to reach each team on time when the sampling methodology changed from Procedure A to B to C. Therefore, a number of the teams in the North simply went ahead with the original Procedure A listing, as against the changing procedures in other regions. In fact, by the time Procedure C was being implemented in Southern States, most teams in the North had already finished the interviews (see sections IVB, C below).

The Southwest sample consisted of three states besides Lagos – Osun, Oyo and Ondo. The fieldwork there was coordinated by Franklin Agbai. Anticipating a tedious and painstaking survey, the team quickly had a review meeting in Ibadan following the conclusion of training. Despite this, the supervisor was continually stretched as he attended to the diverse needs of teams for intervention on gray areas of the work and unexpected experiences with respondents and communities. Teams had to continually invent ingenious means of dealing with challenges, including household members (especially younger ones) asking for tips before allowing the administration of questionnaires, team members being mistaken for government workers on prying missions despite producing identity documents, etc. Fortunately most of the interviewer were young or could pass for students. So where convenient and could ease interaction with households, enumerators who were affiliated with universities in one way or the other would produce identity cards to that effect instead identity card and introduction letter from Zibah Consults.

Owing to the stress of the daily interviews, again it was not practicable to input data from the interviews on a daily basis as hoped. Teams were therefore allowed to wait till the end of the field work to start the data input process. However, team leaders and where possible the supervisor prior to that had to crosscheck the entries in the hard copies each day to ensure that mistakes were corrected while the team was still within reach of the respondents. Teamwork among the members of the region was appreciable. Owing to the high workload, Oyo had to engage one extra interviewer while Lagos engaged two more. They were trained in the field and in the listing and interviewing procedures. It is noteworthy that the work in Lagos was perhaps the most intensive of all the states in the country, given the double sampling of the state, leading to large number of local governments that had to be covered. Consequently, the team in Lagos first concentrated on listing households in the first set of local government areas assigned to them before beginning interviewing. They put in very commendable efforts to cover the work in good time.

The experiences of the different teams provide important lessons for the design and administration of future surveys. As one of the supervisors noted, time pressures led to insufficient work on design issues and experimentation before the actual fieldwork. The result was the need to later make costly changes after the teams were already in the field. This required more work of supervisors to ensure that not only were agreed-upon standards

maintained by every team, but that this was done in a consistent manner. Communication problems led to significant avoidable cancellations and repetitions of procedures in the field. Any deficiencies in the time for adequate planning lead to costly adjustments in the field. This makes it imperative to conduct thorough and detailed preparatory work before going out to the field. Of course, no amount of preparatory work can completely eliminate chance occurrences in the field, but it minimizes them and enhances pre-emptive reaction to them. In particular, challenges came with the changes in sampling methodology. The changes in the sampling procedure, the coverage and selection of sample areas (states, local governments and enumeration areas) when teams were already in the field and had started fieldwork negatively affected the fieldwork. It took time to cascade the new instructions to the teams in the field, including pointing out the differences between the old and new methodologies and explaining the rationale behind the changes.

Dealing with the question of confidentiality and by extension obtaining sensitive information from respondents remains a continual challenge in household surveys of this nature. The Project Team would not say that it found the magic wand but it is proper to assert that understanding cultural sensitivities and winning the confidence of respondents is critical to obtaining meaningful results. The Nigerian survey recorded only one non-response from all teams. This was possible because every supervisor was made to appreciate the need to seek the approval and assistance of traditional local leaders. Consequently, local publicity including town criers backed by the traditional ruler and his chiefs helped secure the cooperation of the respondents. In one instance, the team was treated to a feast by the chief, who then gathered the sample households for interviewing. To a large extent then, not only was the sampling easy, but there were greater grounds for placing faith in the information obtained. On the other hand, in many urban areas, particularly in the south where the traditional governance institutions have been weakened, as in Lagos, Ibadan and other urban areas, teams had to resort to appeals, diplomacy and persuasion to obtain results from sample households.

A significant number of supervisors and interviewers appealed for reduced and more focused questions in the instrument. But the project management team in Nigeria made it clear it did not have the latitude for such significant changes as the project was continent-wide and had to be made comparable to other countries. So, changes were limited to minor adjustments in wording that contextualized the questions for Nigeria and yielded a better understanding of each question. While this procedure sufficed for the present project, the demand for greater clarity in the instrument (which we verified did not simply stem from the desire for reduced work) indicates the necessity of more extensive deliberations with potential participants in fieldwork before commencement of the survey in the future. While participants cannot be allowed to decide the contents of the instrument, it might help to have an outline of targeted outcomes and then have a wider audience make inputs into the nature of questions to yield better results in each country.

B. Modifications to the Initial Sampling Procedure in the Last Stage (Sampling Households)

As the fieldwork progressed, it was found desirable to make adjustments. The pilot study of two local government areas in two states – Enugu and Kogi – had showed the incidence rate of international migration to be about 10 percent (5 households out of 50 interviewed), 64 percent for internal migration (32 households) and 26 percent for non-migrant households (13

households). However, in the course of the actual fieldwork, some early household listings in states in the South (East) indicated that the relative number of households with international migrants was higher than observed in the pilot, with a few extreme cases being in excess of 20 percent. So it was decided that it would be excessive to list as many as 100-150 households to select 34. To address this, the number of households to be listed by teams was reduced to between 50 and 100, and the number to be interviewed pegged at about 20 for each EA, but including specific guidelines on how to oversample households (hhs) with international migrants relative to others, and households with internal migrants relative to those with no migrants. This took into account the expected relative prevalence of the three types of households. Thus it was decided, in a new Procedure B, to select a maximum (to reduce clustering effects, from large clusters) of 10 households with migrants in each sample EA. This meant that if there were fewer than 10 listed, all would be sampled and accepted to be interviewed, while in the few cases expected to have more than 10, the 10 would be selected randomly, keeping the maximum at 10. In stratum 2, the maximum would be fixed at 7, recognizing that most sample communities would have a tenth or more of their households with an out-migrant in the past 9 years to some internal destination. Then for stratum 3, we set a minimum of 3 non-migrant households and a maximum of 12. Thus in an EA with many out-migrant households, the maximum numbers of households with international migrants and internal migrants selected in the sample would be 10 and 7, leaving 3 non-migrant households to sum to 20. If there were a shortfall in the number of households with international migrants, the number in stratum 3 would be increased. For example, if 60 households were listed in an EA, one having an international migrant, 5 having internal migrants, and 54 having non-migrants, then the sample would comprise 1 international migrant household, 5 internal migrant households, and 12 non-migrant households, for a total of 17 households in the EA. If there were 0 international migrant households, 10 internal migrant households, and 60 non-migrant households, the numbers sampled in the three strata would be 0, 7 and 12, for a total of 19. The understanding was that there would usually be a higher number of households with no migrants than internal migrants and almost always a larger number of internal migrants than international migrants leading to higher sampling fractions for the latter in each case. Even with this, the total number of households with internal migrants in the final sample interviewed was still expected to be larger than the number with international migrants because the latter were far rarer. But generally, the difference would be less with this Procedure B than with the earlier Procedure A. Note this meant there would always be a range of 0 to 10 households to be interviewed in stratum 1, 0 to 7 in stratum 2 and 3 to 12 in stratum 3. Where there were no households in strata 1 and 2, the number in stratum 3 could be increased in each EA. This procedure was named Procedure B to distinguish it from the original procedure (termed Procedure A, explained earlier).

But after reducing the number of households to be selected per EA to 20, a new challenge emerged namely; the overall number of questionnaires to be generated given the total number of enumeration areas and the number of households within each enumeration area to be interviewed would yield less than the desired 2000 households. Thus, it was decided to increase the range of households to be interviewed in each household stratum and peg the number of households to be interviewed per EA at a maximum of 25 instead of 20. For households with international migrants, the maximum number was accordingly increased to 12, with the number of internal migrant households increased from 7 to 10. The minimum of 3 was retained for households without migrants, with the maximum increased to 25. This modified procedure was termed Procedure C, to distinguish it from the first and second procedures (A and B) for sampling outlined earlier.

Thus, household sampling was carried out using 3 different procedures, depending on when the EA was covered, in the search for an optimum oversampling process to identify the (relatively rare) sample households with recent out-migrants to international destinations. Procedure A involved sampling about 34 randomly selected households from a list of 100 to 150 households in each sample enumeration area. The number of enumeration areas completed with this depended on the speed of the team in each state (see discussion below in sections IV B, C). Most enumeration areas in the North were sampled using procedure A: all in Bauchi, Kano, Katsina, Niger and Yobe. Only Zamfara did not use A, and used B instead. Under procedures B and C (see discussion below in section IV B, C), between 50 and 100 households were listed in each sample EA, from which about 20-25 were selected by team leaders in the field using the supervisor sampling sheets.

In addition to the adjustments in household listing and sampling procedures, the number of local government areas in sample states was increased by 50 percent across the two strata of (high and low migrant) states.[4] This had to be done in a manner that maintained the ratio of 2:1 in favour of households in the high stratum states to ensure oversampling of this group. This implied adding one local government area in each of the sample states in the South and adding one local government area for three of the selected six states in the North. For the latter, consideration was given to the existing coverage of the three geopolitical zones in the region i.e., North West, North Central and North East. Given its huge population, Kano State in the North West was automatically selected to have an additional Local Government from among Kano, Katsina and Zamfara in the zone. Niger state was the only state in the North West included in the sample, so was also selected to have an additional LG. Finally, Yobe was selected from between Yobe and Bauchi from the North East.

For the south, it was agreed that as many states as budgetary constraints could allow for should have an additional LG. Eventually, with prudent resource management, it was possible to add one LG to every state in the sample in the high stratum South (including 2 in Lagos to reflect its double sampling). Except where impracticable, the new set of LGs was again to be drawn from the original NBS list, but this was not possible on two occasions. The first was in Ebonyi State, where one of the communities selected was involved in communal clashes with a neighbouring community. The second was in Delta state where one of the selected communities was prone to activities of militants, crisscrossed by creeks, and had difficult swampy terrain. On both occasions, the supervisor for the region was allowed to randomly select another LG which did not necessarily have to be in the NBS list but which increased equitable coverage across the senatorial zones of the state. The stages of sampling described above as well as the adjustments gave rise to the states, local governments and enumeration areas shown in Table 1.

C. Listing and Sampling in the States

As noted in the previous section, there were eventually three procedures for listing and sampling for the teams. The points of communication of these changes for the teams determined the relative number listed and sampled in each enumeration area. As at the time of the transition from Procedure A to Procedure B, a number of teams had completed a certain proportion of interviews, while other teams, which undertook to complete all the listings first, with the intent to sample households and begin interviews afterwards, were still able to adapt to Procedure B to sample households. Teams in the second group were

---

[4] It was necessary to increase the sample of LGAs to compensate for the smaller mean number of households expected in Procedures B and C (20-25—see below) compared to Procedure A (34), which had been assumed in selecting the original sample size of 60 EAs.

instructed to not bother re-listing, but instead to adjust the numbers in the three lists to be sampled from. For those teams that were both listing and interviewing, the speed of each team determined the point at which it was instructed to change household listing and sampling procedures. The same issues arose at the time of the second change, from Procedure B to C, though the adaptations necessary were much less since the listing aspects were identical.

Within the ambit of each procedure, teams were given a range of number of households to list and sample. For example for all the Procedures (A, B and C), there was a range of about 50 households between the maximum 150 and 100 or 100 and 50 that enumerators could list. Likewise, a recommended maximum number was given for sampling. Ultimately, the number listed by each team depended roughly on the proportion of households with international migrants. For areas with low incidence of international migrants, more households were listed, to increase the probability of finding the population of interest, i.e., households with international and internal migrants (where the latter was also rare). In places where out-migration rate was high, only around 50 households were listed since the population of interest was more easily available for sampling. The ultimate number listed or sampled by each team depended in part therefore on the prevalence of international migrants being encountered during listing.

As such, in nearly all states in the south, listing and sampling proceeded with a combination of procedures. For example, Abia, Delta, Ebonyi, Imo, Ondo, Osun, Oyo and Rivers worked with a combination of procedures B and C, since the first change from procedure A to B happened while they were still listing. Thereafter, Procedure C was introduced when the additional Local Governments were added. In contrast, all the Northern States, where the LGs and EAs to be covered were fewer in number, the listing had proceeded quickly with procedure A partly due to the very low proportion of households with migrants (international and internal). All states in the North (except Zamfara) therefore used procedure A, which involved listing more households. The team in Lagos had also started with Procedure A in the first set of EAs, and was still doing listings when the announcement for the change from procedure B to C was made. So the sampling switched from Procedure A directly to Procedure C, leaving them without any EA sampled with procedure B. Edo state was also listing when the first change was made and therefore finished up its listing within the framework of the newly adopted procedure (B).

The above scenarios produced a wide range of listing and sampling procedures across states, but all consistent with the specified methodologies. The number of households listed ranged from 660 in each of Ondo and Osun (Lagos had 1084, but this is because it was sampled twice) to 102 in Zamfara (having only one Local Government and two EAs). Likewise the number interviewed ranged from 150 in Akwa Ibom to 70 in Katsina and 49 in Zamfara. The average proportion of households interviewed relative to those listed is shown in the last column as 30 percent. The states and final samples of households are presented in Table 2.

Table 2: Sample Listing and Interviewed Households

State	LGs	EA	Procedures Used	Listed	Int'l	Internal	Non-Migrant	Interviewed	Int'l	Internal	Non-migrant	Proportion of List Sampled (%)
Abia	3	6	High	B, C	524	59	320	145	132	46	52	34	25.19
Akwa Ibom	3	6	High	C	385	42	190	153	150	36	74	40	38.96
Bauchi	1	2	Low	A	203	0	8	195	67	0	7	60	33.00
Delta	3	6	High	B, C	276	56	168	52	120	48	56	16	43.48
Ebonyi	3	6	High	B, C	508	35	349	124	142	28	63	51	27.95
Edo	3	6	High	B	479	87	240	152	124	29	62	33	25.89
Imo	3	6	High	B, C	649	83	324	242	129	63	45	21	19.88
Kano	2	4	Low	A	400	4	161	235	134	3	66	65	33.50
Katsina	1	2	Low	A	219	0	14	205	70	0	14	56	31.96
Lagos	6	12	High	A, C	1084	124	440	520	332	115	148	69	30.63
Niger	2	4	Low	A	417	0	47	370	136	0	40	96	32.61
Ondo	3	6	High	B, C	660	59	312	289	130	56	43	31	19.70
Osun	3	6	High	B, C	660	35	172	453	140	27	61	52	21.21
Oyo	3	6	High	B, C	598	179	274	145	128	46	42	40	21.40
Rivers	3	6	High	B, C	502	73	118	311	132	66	46	20	26.29
Yobe	2	4	Low	A	409	0	60	349	136	0	41	95	33.25
Zamfara	1	2	Low	B	102	0	15	87	49	0	15	34	48.04
Total	45	90			8075	836	3212	4027	2251	563	875	813	30.17

Note that overall, the proportion of households listed with international migrants that were selected into the sample was about 2/3, compared to about 1/3 of internal migrant and 1/5 of non-migrant households. This shows the value of oversampling at the last stage (of two-phase sampling).

It is also useful to take a 6-region perspective to the listing and sampling. For this purpose, we use a slightly different classification from the usual geopolitical split, by merging the North East and North Central, since both are low migration states and have a combined total of only 3 states in the sample. In turn, we deal with Lagos as one region, given that its total of 6 Local Government Areas was higher than the number of any of the regions in the North (4 in the North West, 3 in the North East and 2 in the North Central). All states in the North contributed a combined 9 Local Governments in the sample, compared to 9 in each of the South East and South West regions (excluding Lagos) and 12 in the South South.

Of the 8,075 households listed in the project, 1,918 or 24 percent of the total came from the three states in the Southwest. This was followed by the Southeast with 1681 or 21 percent of households. The South South follows with 1642 households, representing 20 percent. The Northeast/North Central zone posted 1029 alongside Lagos with 1084, both having approximately 13 percent of the listing apiece. The Northwest region has the least listing, with 721 households or 9 percent. We summarize this in table 3 and present the proportional distribution in figure 1.

**Table 3: Regional Distribution of Listed and Sampled Households**

| Region | Proportion Sampled | EA | Listed | | | | Interviewed | | | |
			Int'l	Internal	Non-Migrant	Total	Int'l	Internal	Non-migrant	Total
North West	37.83	8	4	190	527	721	3	95	155	**253**
North East/Central	32.96	10	0	115	914	1029	0	88	251	**339**
South East	24.34	18	177	993	511	1681	137	160	106	**403**
South South	33.66	24	258	716	668	1642	179	238	109	**526**
South West	20.77	18	273	758	887	1918	129	146	123	**398**
Lagos	30.63	12	124	440	520	1084	115	148	69	**332**
Total/Ave	**30.17**	**90**	**836**	**3212**	**4027**	**8075**	**563**	**875**	**813**	**2251**

The composition of international, internal and non-migrant households in the listings according to zones presents even more varied representation. With 273 aggregate and a 33 percent share, the Southwest has the highest number of households with international migrants listed, followed by the South South with 258 representing 31 percent and the South East with 177 cases and a 21 percent share. Lagos state contributes the remaining 124 households or 15 percent share. The listing for the entire Northern region (Northwest, Northeast and North Central) combined shows only 4 households with international migrants in Kano state and none from the other states in the region. Figure 2 shows the distribution of households with international migrants as contained in the summary listing sheet for all states according to their respective regions.

Fig. 1: Listing of Households by Regions

Total Listing by Regions
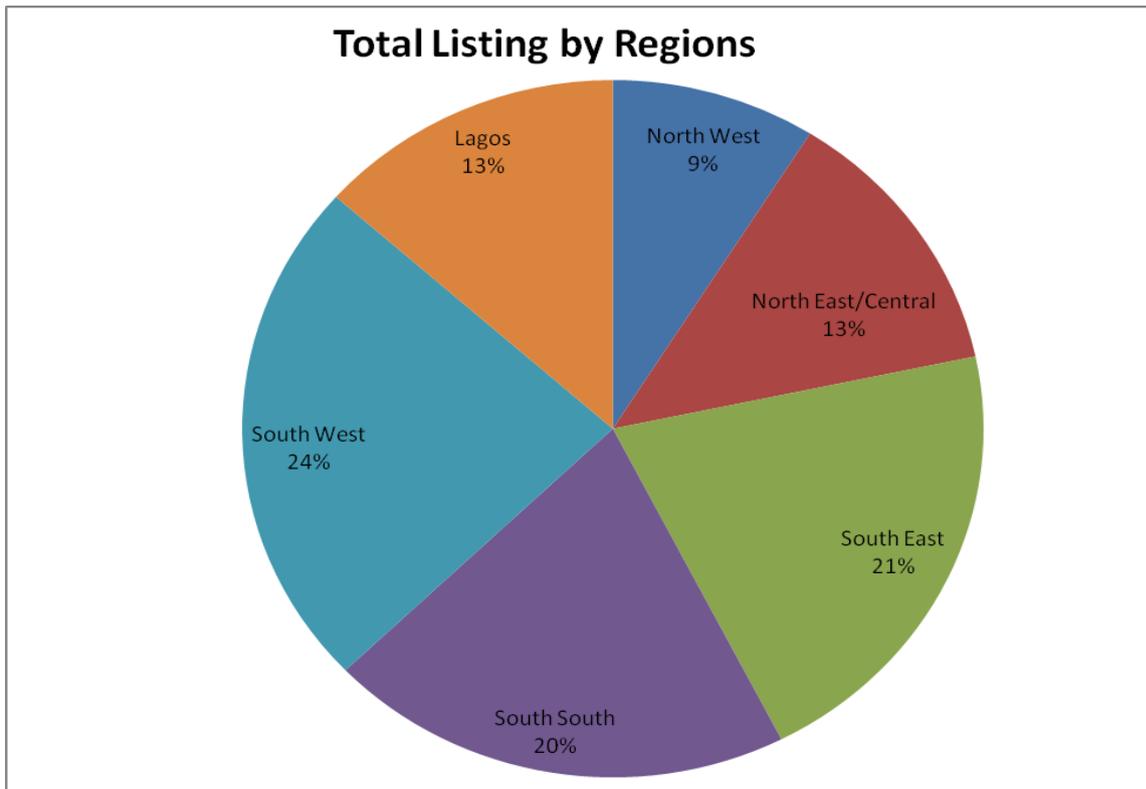
Figure 2: Listing of Households with International Migrants by Regions

Int'l Migrant HHs in Sample by Regions
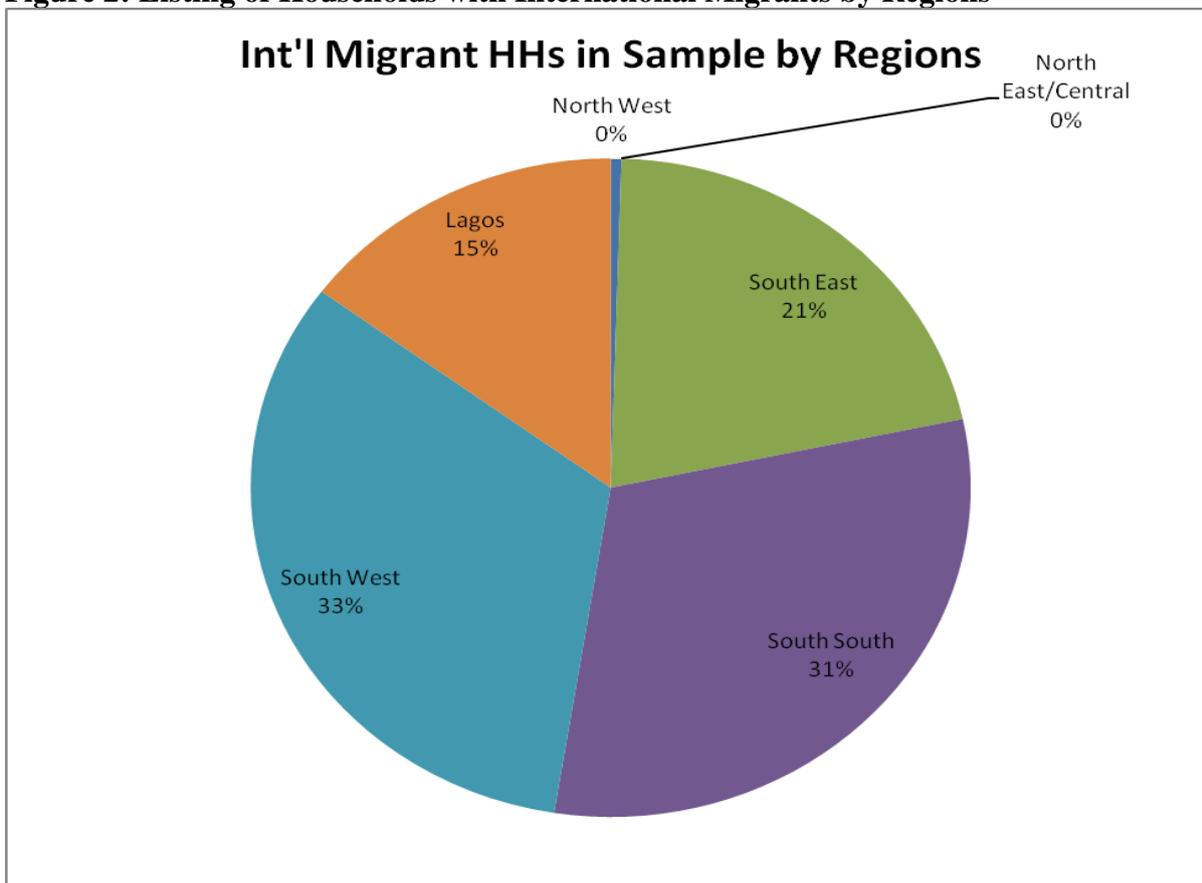
The three regions in the North did have some more households with internal migrants, with the Northwest having 190 households or 6 percent of all households with internal migrants in

the listing. But the Northeast had a much lower share, with 115 households or 3 percent,which gives the entire Northern region 305 households or only 9 percent of the households with internal migrants in the listing. This figure is lower than that of Lagos State alone, which had a total of 440 households with internal migrants or 14 percent of the total. The South East has the highest number of internal migrants, with no less than 993 or 31 percent of all households with internal migrants coming from the zone. It was followed by the Southwest with 758 households or 24 percent of all hosueholds with internal migrants in the list. The South South has 716 or 22 percent of the households with internal migrants in the list. We show the share of each region in composition of households with internal migrants in Figure 3.

Figure 3: Listing of Households with Internal Migrants by Regions
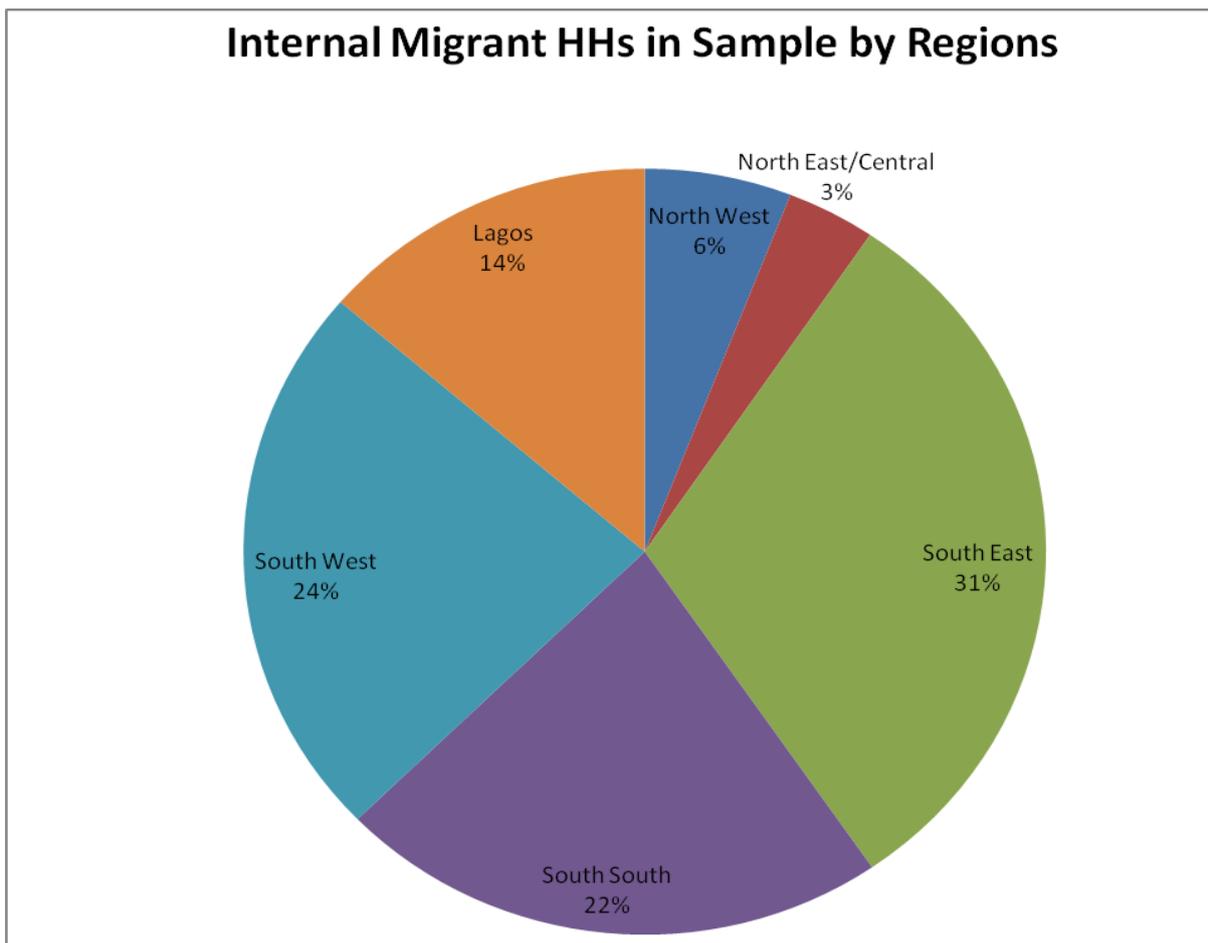
As expected, and given their very low relative shares in both internatinal and internal migrant households, the states in the North have a greater share of non-migrant households. The total listed non-migrant households in the Northwest is 527 (13 percent of total listed non-migrant households) while those from the North East and North Central together constitute 914 or 23 percent of listed non-migrant households. The combined total from the Northern region is 1441 non-migrant households, or 36 percent of the total. The Southwest closely follows the Northeast/North Central region with an aggregate of 887 non-migrant households (or 22 percent of all non-migrant households in the sample list). Other numbers are seen in the table, and presented in Figure 4.

Figure 4: Listing of Non-Migrant Households by Regions
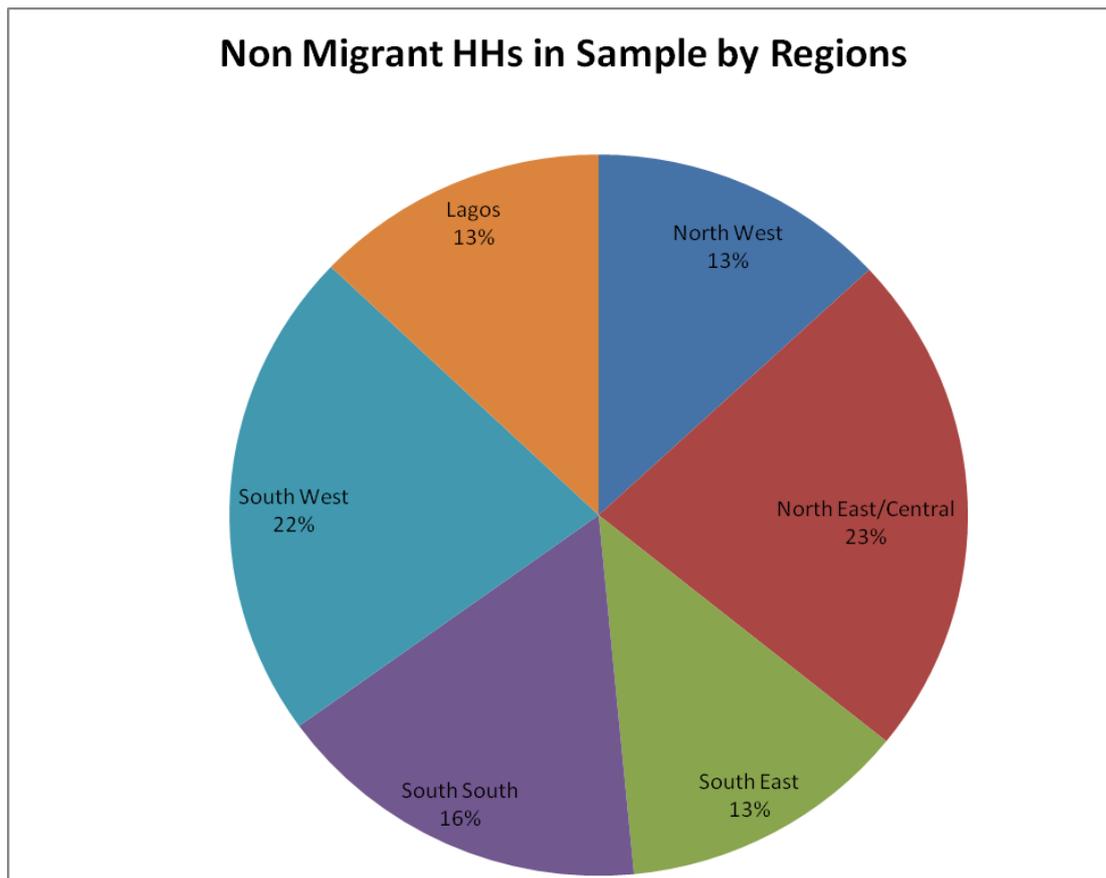
V.   Sample Weighting and Data Quality Issues

Weighting the data is critical in this project given that the sample selected is *not* an *epsem* sample, in which all elements (households) in the sample had an equal probability of being selected. Rather, the sample was explicitly designed to oversample areas and households with international migrants. Therefore, it is not a self-weighting sample in which the (unweighted) sample mean of the values for sample households is a reliable estimate of the population mean. Instead, the sampling procedures used aimed at selecting a sample that would have sufficient numbers of households with recent migrants to international destinations, which are relatively rare elements in the large population of Nigeria (or most any national population). This required the use of special sampling procedures (see Bilsborrow et al., 1997; Groenewold and Bilsborrow, 2008) to ensure that the actual sample of households interviewed included sufficient households with migrants, i.e., much larger than their share in the national population.

Overall, the sample selection involved four-stages, the selection of states, the selection of Local Government Areas, the selection of Enumeration Areas, and the Selection of Households. Though they have been severally referred to and described above, we briefly review each selection stage as follows. The Selection of States was the first stage, which required stratifying all states in the country according to the expected proportion of households in each state with international migrants. Due to the lack of access to data on migration from the 2006 census (the most recent), expert opinion and available literature were relied upon to classify the 36 states and the Federal Capital Territory into high and low

migration strata. Eighteen states were classified in the high migration stratum (16 in the South, 2 in the North), and 19 states were classified in the low migration stratum (18 in the North, 1 in the South). The proportion of the population in the high migration stratum to be sampled was then taken to be double that of the low stratum. This yielded a sample of 12 states out of the 18 in the high stratum and 6 states out of the 19 in the low stratum, with an approximate relative weight of 3/1 for all households in the low stratum compared to 3/2 for households in the high stratum, or double the weight for households in the low compared to the high stratum. To be more precise, the total population of the high stratum states in the 2006 census was 67.04 million, compared to 72.95 million in the 19 states in the low stratum. Therefore, stage I weights for households in the two strata are $67.04/12 = 5.5867$ in the high stratum compared to $72.95/6 = 15.4917$ in the low stratum. This approach achieves an additional benefit in that it stratifies the entire country by population so that the population of the sample can represent the population of the entire country

The second stage of the sample was the selection of Local Government Areas (LGAs), which constitute the next lower administrative units in states. To further concentrate the sample of areas among those expected to have a higher proportion of international migrants; it was decided to select two LGAs from each sample state in the high stratum and one LGA from each in the low stratum. As explained in Section IVB, it was realized that the initial number of LGAs was insufficient to produce the total sample size desired, which was corrected for by (1) increasing the number of LGAs in all sample areas by 50 percent, and (2) increasing the number of households to be selected at the last stage from 20 to 25 (moving from Procedure B to Procedure C, explained earlier). This translated to selecting 3 LGAs instead of 2 in each high stratum state sampled and 1.5 instead of 1 in each low stratum state. The final result was that All states in the South had three LGAs in the sample (Lagos had six as it was sampled twice). Kano, Niger and Yobe were selected from the low stratum states to have 2 Local Governments each instead of 1.

At the Local Government sampling stage, it was considered important to explicitly incorporate the rural/urban divide of the sample in order to better represent these two groups of the population. This was particularly important given that it was not incorporated into the selection of states and in any case, there is no exclusively rural or exclusively urban state in the country; each state has a share of both. Ordinarily, it could make sense to divide the LGAs into rural and urban, and then purposively sample one of each from each sample state. But according to Oluwasola (2007) and the United Nations (2008), Nigeria is one of the most rapidly urbanizing countries in the world. In 2006, the urban population was estimated to be 46 percent of the total, up from 11 percent in 1952. This implies that about 65 million of the country's 140 million lived in urban centres in 2006; so the urban population was estimated to be about half by the time of the survey in late 2009. Given the above, a random sample of Local Governments Areas would produce approximately the same statistical result as a purposive sampling if the target was to have equal numbers of urban and rural LGAs (discussed also above).

To weight the LGAs, the number selected in each sample state as a proportion of the total number in the state indicates the probability of selection, so the inverse is the LGA weight for the state. Table 4 below indicates the weights for each state based on its stratum, the total number of LGAs in each sample state, the number in the sample, and the weight being the inverse of the probability of selection. Thus every selected household in the sample LGA will have the weight indicated attached to it, in the final computation of household weights. The consequent state and LGA weights are shown in Table 4 below.

Table 4: Sample States and LGAs Weights

	State	Stratum	State Weights	No of LGAs	No. in Sample	LGA Weight
	NORTH					
1.	Abia	High	5.587609	18	3	6
2.	Akwa Ibom	High	5.587609	31	3	10.3333
3.	Bauchi	Low	12.15870567	19	1	19
4.	Delta	High	5.587609	25	3	8.3333
5.	Ebonyi	High	5.587609	12	3	4
6.	Edo	High	5.587609	15	3	5
7.	Imo	High	5.587609	27	3	9
8.	Kano	Low	12.15870567	45	2	22.5
9.	Katsina	Low	12.15870567	34	1	34
10.	Lagos	High	5.587609	20	6	3.3333
11.	Niger	Low	12.15870567	25	2	12.5
12.	Ondo	High	5.587609	19	3	6.3333
13.	Osun	High	5.587609	30	3	10
14.	Oyo	High	5.587609	32	3	10.6667
15.	Rivers	High	5.587609	23	3	7.6667
16.	Yobe	Low	12.15870567	17	2	8.5
17.	Zamfara	Low	12.15870567	15	1	15

The third stage was the selection of Enumeration Areas from LGAs. The total number of EAs in Nigeria in the 2006 census is about 23,000, with the number per state varying from 180 to 1350 and the number in sample states from 360 to 1350. The reason for this variation is that the number of LGAs per state varies so much though the number of EAs per LGA remains 30. So it should have been desirable to obtain the census population of each sample LGA and of each EA in each sample LGA to know what the population in the sample EA represents. These data were also not available from the National Bureau of Statistics; only information on the population of each LGA was available. We thus assumed that the population of each EA in the sample LGA was approximately the same. Thus Given that 2 EAs were selected out of 30 EAs in every sample LGA, the probability of the selection of an EA was 2/30, and the weight for all sample EAs in the survey is straightforward, i.e., the inverse of 30/2 or 0.06666.

The last stage was the selection of households within each sample EA. As has severally been referred to in this report, the survey involved three distinct phases of selection of households from EAs, referred to as Procedures A, B and C. Procedure A involved listing about 150 households in an EA, identifying their migration status, and selecting randomly a total of 34. The instruction given at that time was that any sample household which refused the interview or for any other reason did not provide the data requested should be replaced by another household selected randomly from the list guaranteeing 34 completed households per EA.

However, there was no actual replacement during the fieldwork using procedure A. Since quotas of a maximum of 12 households with international migrants, 12 with internal migrants, and 10 non-migrants were to be selected, with the sum being 34, the weights were to be computed as follows for those EAs in which Procedure A was used (mainly in the North). So if X1, X2 and X3 are the actual numbers of households listed in the three strata, and n_1, n_2 and n_3 are the numbers actually sampled (following the procedures described in III B above), then the weight for each international migrant household selected in the EA is $X1/n_1$, while the weights for each internal migrant and non-migrant household are respectively $X2/n_2$ and $X3/n_3$.

Procedure B involved listing 50 to 100 households in the sample EA, and recording the numbers in each stratum. In the first step, up to 10 households with one or more (recent) *international* migrants were selected. Where the total number of households with international migrants in the list was less than or equal to 10, all were selected, meaning each of those households had the same 100 percent chance of being selected. Where the number selected was less than the total number of households with international migrants listed in the EA, then the probability of selection was 10/X1, where X1 is the total number listed, and the weight for each household in stratum 1 is X1/10. In the second step in Procedure B, households with *internal* migrants were selected with the probability calculated as the ratio of the number selected to the total number in the stratum, allowing a maximum of 7. Again, if there were more households listed than 7, say Z, the probability of selection of each is 7 divided by that number, and the weight is the inverse, or Z/7 or Z/n_2. The last step was to sample *non-migrant* households in the third stratum. For any numbers fewer than 10 in stratum 1 or fewer than 7 in stratum 2, there would be a corresponding increase in the sample size selected from stratum 3. This number was determined before drawing the sample, but would always be at least 3 since a maximum total sample size from each EA was set at 20, meaning that even with a maximum selected in strata 1 and 2 (17), three non-migrant households would be selected. Again, the weight is $X3/n_3$, as before, but following the supervisor sampling sheet for Procedure B. Finally, Procedure C is similar in methodology to Procedure B except that the total number or "sample take" per EA was increased from 20 to 25 – with the numbers in the sample in strata 1 and 2 (international and internal migrant households) allowed to increase from 10 to 12 and 7 to 10, respectively. This left the residual for non-migrant households at the same minimum of 3 for stratum 3, but a maximum of 20. Note that in Procedures B and C, there was no replacement of non-responding sample households, so the weights in all cases in both procedures B and C (and as is normal in surveys where there is *no replacement* of non-responding sample households) can appropriately take into account that some households selected in the sample from the list may refuse, not be at home, etc.

To bring all the weights together from the four-stage sampling procedure, we multiplied the four weights for sample state, sample LGAs within the state, sample EAs within the LGA, and the appropriate stratum weight of the household in its sample EA. So for a particular household H in a state j denoted by H_j, the probability of selection would be the consolidated probability of the migration stratum in which the state it falls into multiplied by the weight of the LGA within the State multiplied by the weight of the Enumeration area within the LGA multiplied by the household weight based on its selection within the stratum of households that it falls into in its EA. So if weights for selecting states in stage 1 in the high and low strata are indicated by W^1_m where the subscript m=1,2 stands for either high or low strata; weights for LGAs in each sample state in stage 2 are given by W^2_L with subscript L indicating the sample LGA in the state; and weights for selecting EAs from sample LGAs in

stage 3 are designated by W^3_e with subscript e representing 30/2. Finally, weights for households in the last stage are given by W^4_S, with the subscript S representing any one of the three household strata (international, internal and non-migrant households). The total weight W applied to a household (or individual migrant in a household) in Stratum S and EA designated by e, within Local Government L in State m, is given by:

$$W_{mLeS} = W^1_m * W^2_L * W^3_e * W^4_S$$

This procedure was adopted to calculate the weights for all sample households, with the weights for sample households in the three strata in each sample EA in each sample LGA in each sample state presented in Annex A.

Given all the precautions, diligence and effort put into the data collection, it is anticipated that the data represent well the population of Nigeria, and that the data collected in the survey should be of fairly high quality, despite the difficult circumstances. Nevertheless, there are not only sampling errors, which are part of any sample survey, but also non-sampling errors due to occasional errors associated with respondents providing incorrect information, interviewer errors reading/explaining questions or recording responses, and data entry errors, but we tried to minimize these by recruiting well-educated, experienced interviewers and laboriously checking data entered and correcting data entry errors found.

VI. Conclusion

The entire data generation and processing for the migration study in Nigeria was a very challenging--and equally very rewarding--experience. The lessons are many and the experience immensely exciting. Each of us involved in the Nigeria Migration Survey project learned much as we threw in our best efforts to collect data that reflect the migration reality on the ground, which will also be relevant to policy formulation. The Zibah team believes the data are the best that exist on migration in Nigeria. It has been our privilege to be part of the process of producing the first generation of quality data on migration across developing countries in Africa through the World Bank project, yielding data that will make possible a better understanding of the growing impact of migration and remittances on development. It is our hope that the data will serve as useful inputs into the policy development process to enhance the feedback of human movement on development around the world and particularly in developing countries.

References

Agu, Chukwuma, 2009, "The Remittance Service Industry in Nigeria: An Assessment" Background Paper for Africa Migration Project, Migration and Remittances Team, Development Prospects Group, the World Bank, Washington D.C.

Bilsborrow Richard E, Graeme Hugo, A. Oberai, and Hania Zlotnik, 1997.. *International Migration Statistics: Guidelines for Improving Data Collection Systems*. Geneva: International Labour Office, 441 pp.

Groenewold, George, & Richard E. Bilsborrow, 2008. "Design of Samples for International Migration Surveys: methodological considerations and lessons learned from a multi-country study in Africa and Europe", in *International Migration in Europe: New Trends and Methods of Analysis, Bonifazi, C. et al. (Eds),* Amsterdam University Press: 293–312.

Hernández-Coss, Raúl, and Chinyere Egwuagu Bun (2007), "The UK–Nigeria Remittance Corridor: Challenges of Embracing Formal Transfer Systems in A Dual Financial Environment" World Bank Working Paper 92: Washington, D.C.: The World Bank

Oluwasola, Oluwemimo, 2007 "Social Systems, Institutions and Structures: Urbanization, Poverty and Changing Quality of Life" Paper Presented at the Training Session of the Foundation for Environmental Development and Education in Nigeria, February 22, 2007. Obafemi Awolowo University, Ile-Ife, Nigeria.

United Nations Department of Economic and Social Affairs, 2008, "An overview of urbanization, internal migration, population distribution and development in the world" Paper Prepared for the United Nations Expert Group Meeting on Population Distribution, Urbanization, Internal Migration and Development, New York, January 21-23, 2008 United Nations Population Division (DESA) (UN/POP/EGM-URB/2008/01)

Annex A: Enumeration Areas and Household Weights in the Sample

State	LGA	EA Names	Listed	Inter viewed	Weights Int'l	Internal	Non-Migrant
Abia			**909**	**282**			
	Umuahia North	Umunkaru	102	22	1.600	7.111	7.333
		Utali Ofeme	94	20	1.000	5.250	6.750
	Bende	Ndielu	80	20	1.444	5.714	6.750
		Amaba	80	20	1.333	7.333	3.500
	Ikwuano	Umudike	84	25	1.100	6.444	2.500
		Ndoro	84	25	1.000	5.667	2.889
	Etinan	Ekom Iman	54	25	2.000	2.000	2.667
		Ikot Obio Inyang	53	25	1.000	2.273	2.273
	Onna	Abat	83	25	1.000	4.500	5.600
		Ikot Ebekpo	85	25	1.143	4.000	12.333
	Uyo	Afaha Offot/ Bassey Udoeka	55	25	1.200	1.714	4.167
		Effiat Offot/ Esuene St	55	25	1.500	2.143	2.444
Bauchi			**203**	**67**			
	Ningi	Wuga	100	34	0.000	1.143	3.407
		Jangachi	103	33	0.000	0.000	3.121
Delta			**276**	**120**			
	Isoko North	Owhelogbo	40	20	1.000	2.364	2.000
		Ozoro	40	20	1.400	1.846	4.500
	Isoko South	Emore Rd 1	50	20	1.111	3.889	2.500
		Emore Rd 2	61	20	1.100	4.625	6.500
	Warri S/West	Isaba	40	20	1.364	3.667	1.000
		Ogbe-Ijoh	45	20	1.000	2.667	6.000
Ebonyi			**508**	**142**			
	Afikpo North	Egeburu	106	20	1.400	7.364	4.500
		Ugwu-Egu	101	19	1.000	7.900	2.857
	Ezza North	Oriuzo	100	24	1.000	5.889	3.286
		Ohagolode	85	25	1.000	7.556	1.071
	Ohaozara	Onuogo	56	20	1.000	4.000	2.500
		Amechi Okposi	60	34	1.385	2.133	1.667
Edo			**479**	**124**			
	Esan North	Efandion	80	20	3.000	2.700	7.600
		Uromi	80	20	6.500	2.182	6.143
	Ikpoba Okha	Ihinwinhin	80	21	2.700	7.000	3.000
		Ugbekun	80	20	3.500	4.556	3.600
	Owan East	Ikhin	80	20	1.500	4.750	3.333
		Warrake	79	23	2.000	3.733	3.750
Imo			**649**	**129**			
	Owerri Municipal	Umuodu	119	20	1.125	9.625	8.250
		Umuoyima	119	20	2.200	7.857	14.000
	Ezinihite	Oparachi	78	20	1.167	7.167	10.500
		Umuekwene	112	20	1.500	7.000	16.000
	Owerri North	Umuodu Awaka	119	25	1.000	7.875	8.800

State	LGA	EA Names	Listed	Inter viewed	Int'l	Internal	Non-Migrant
					Weights		
		Umudagu	102	24	1.000	4.111	13.500
Kano			**400**	**134**			
	Minjibir	Minjibir	100	40	1.000	1.483	5.600
		Kadani	100	27	2.000	2.000	5.833
	Gaya	Gaya	100	34	1.000	4.400	2.391
		Shagogo	100	33	0.000	3.538	2.700
Katsina			**219**	**70**			
	Kaita	Sabongari	109	34	0.000	1.000	3.586
		Gande	110	36	0.000	1.000	3.741
Lagos			**1,084**	**332**			
	Ajeromi/Ifelodun	Ifelodun	110	33	1.333	3.000	5.333
		Immam	110	34	1.500	3.133	5.333
	Mushin	Cash Street	102	33	1.000	3.333	4.778
		Cardoso	110	34	1.000	2.389	6.100
	Ikeja	Adedoyin	100	25	1.000	4.889	11.000
		Otunba	100	25	1.000	4.000	11.500
	Badagry	Ajido	102	21	1.000	5.000	13.250
		Iyesi	110	28	1.000	3.917	12.750
	Oshodi/Isolo	Durowoju	60	25	1.000	2.083	6.500
		Alfa-Nda	60	25	1.000	1.000	9.750
	Surulere	Fadairo	60	25	1.154	1.250	8.750
		Ogunmuyiwa	60	24	1.000	2.556	6.500
Niger			**417**	**136**			
	Bida	Mamalue	108	34	0.000	2.600	3.276
		Abdullahi	105	34	0.000	0.750	4.364
	Chanchaga	Kissi II	102	34	0.000	1.167	4.000
		Wakili Kalma	102	34	0.000	1.000	3.957
Ondo			**660**	**130**			
	Akoko S/East	Isua	110	20	1.200	7.833	12.750
		Epinmi	110	20	1.167	7.250	7.500
	Idanre	Owoseeni	110	20	1.000	7.167	14.250
		Akinala Alade	110	20	1.000	9.333	11.000
	Akure South	Araromi	110	22	1.000	9.000	5.571
		His Mercy Plaza	110	28	1.000	4.500	8.833
Osun			**660**	**140**			
	Ife East	Mojeed	110	20	0.000	4.385	7.571
		Sunboye	110	20	0.000	3.200	12.400
	Boluwaduro	Otan-Ayegbaju	110	25	2.250	1.000	7.154
		Eripa	110	25	1.000	1.000	8.727
	Ifelodun	Apanpa	110	25	1.143	1.167	7.917
		Alamole	110	25	1.182	4.300	13.500
Oyo			**598**	**128**			
	Ibadan North	Customs Qtrs	100	20	2.909	9.800	4.750
		Alhaji Salawu	98	23	6.833	4.300	2.000
	Ibadan S/West	Best Foundation	100	20	2.250	7.429	6.000
		Chief Eniraytan	100	20	7.500	14.500	1.000
	Ibadan N/East	Halleluyah	100	20	4.000	5.429	5.667
		Mr Ajaigbe	100	25	3.000	3.778	6.000
Rivers			**502**	**132**			

State	LGA	EA Names	Listed	Inter viewed	Weights		
					Int'l	Internal	Non-Migrant
	PHC City	Rumuwoji	102	20	1.500	5.286	16.667
		Amadi Ama	80	20	1.000	2.143	18.333
	Oyigbo	Oyigbo	80	20	1.000	1.571	19.667
		Afam Nta	80	20	1.000	3.429	15.333
	Obio/Akpor	Rumuokwuta	80	27	1.154	2.444	8.600
		Rumuigbo	80	25	1.000	1.000	19.333
Yobe			**409**	**136**			
	Gujuba	Gujba	102	34	0.000	3.400	2.931
		Kotorko	105	34	0.000	1.350	5.571
	Gulani	Diyi	100	34	0.000	1.000	3.538
		Yango	102	34	0.000	1.000	3.615
Zamfara			**102**	**49**			
	Bungudu	Baichin Yaman	51	24	0.000	1.000	2.800
		Samawa/Chediya	51	25	0.000	1.000	2.368
			8,075	2,251			