

Annex 1

Sampling Plan for Studies of Title II Development Assistance Programs in Guatemala, Niger and Uganda

Background

In accordance with the evaluation policy of the U.S. Agency for International Development (USAID), Food for Peace (FFP) has contracted with ICF International to conduct a baseline study in Guatemala, Niger, and Uganda for new Title II program awards (July 2012) in these countries. The quantitative component of the baseline survey will be standardized across the participating countries to permit comparative analysis and will collect data for 20 FFP indicators as described in the USAID FFP *Standard Indicator Handbook*. These indicators are related to food access; children's nutritional status and feeding practices; women's nutritional status and dietary diversity; water, sanitation, and hygiene; agricultural practices; and measurements of poverty. In addition to the required FFP indicators, the quantitative survey will also include a small set of program-specific indicators identified by the Title II implementing partners as key measures for their individual programs. The survey design for the quantitative baseline survey will be described in detail in the following document. Most of the details of the survey design were decided upon at a joint meeting with the ICF International in October, 2012. See Appendix A for the minutes of that meeting.

Survey Research Design

These baseline surveys will serve as the first phase of a pre-post survey cycle with the second phase being conducted at the end of the five-year Title II program. Thus, the primary objective of the baseline surveys will be to assess the status of the FFP and program indicators prior to program implementation. The baseline measurements will then be used to calculate change in these indicators (and to undertake a statistical test of differences in the indicators) at completion of the five-year Title II cycle when the same survey will be conducted again in the program areas. This pre-post design will allow the measurement of change in indicators between the baseline and final evaluation; but will not allow statements about attribution or causation to be made.

The baseline surveys will be designed as population-based surveys in the villages/communities selected by the Title II implementing partners in the designated geographic regions of operation. Thus, the sampling frame for each country will only include villages/communities in the geographic regions where the Title II partners are implementing their programs, and will exclude villages/communities where programs are not active. From this frame, a representative sample of villages/communities will be drawn for each Title II partner within each country. Within each sampled community, a representative sample of households and individuals (that includes both beneficiaries and non-beneficiaries) will then be drawn.

Sampling Frame

The sampling frames for each country will be constructed from lists of communities/villages provided by the Title II partners and complemented with census-level household and population

Annex 1

Sampling Plan for Studies of Title II Development Assistance Programs in Guatemala, Niger and Uganda

information in order to assign a measure of size to each selected community. The last available census level information for the geographic regions in each country at the lowest enumeration level will be used. Since the most current census data available in all of these countries is ten or more years old, it is expected that household and population counts will have changed and that newly formed communities will not be represented. ICF will work with the Title II partners to add missing communities and match census-level data with their implementation communities in order to obtain the most up-to-date size information. Appendix B provides summary counts of the implementation communities for each country and program.

Sample Design

Given the availability of size measures for each selected community, cluster sampling with a method that approximates PPS (probability proportional to size) will be used to select communities for each Title II program (two in Guatemala, two in Uganda, and three in Niger). The sample size for each program will be determined based on the selection of one FFP indicator. At the sampling meeting held with FANTA in October 2012, it was agreed that stunting will be used as the primary indicator for deriving sample size estimates since it is a key measure for food insecurity and will provide enough households to measure desired change levels for most other indicators. Additionally, some criteria for sample size calculations were adjusted from ICF's original proposal based on feedback from FANTA. Assumptions for updated sample size calculations for each Title II program are as follows:

- design effect of 2,
- confidence level of 95%,
- power level of 80%,
- expected change in stunting over the life of the program of 6 percentage points,
- use of the Stukel/Deitchler Inflation and Deflation Factors to determine the appropriate number of households (with children aged 0-59 months) to select, as described in the FANTA Sampling Guide Addendum, and
- inflation of the sample size of households by 10% to account for anticipated household nonresponse;

The formula used for deriving sample size is based on a statistical test of the difference of proportions (or prevalence) for an indicator (e.g., from baseline to final evaluation), controlling for inferential error as described in Appendix 1 of the Addendum to FANTA Sampling Guide (March 2012). The table below provides the target sample sizes for each Title II partner program in each country using currently available estimates for the prevalence of stunting and household size in each country. Use of the above assumptions and the revised formula did not significantly alter the sample size calculations provided in ICF's original proposal and, therefore, have no significant cost implications.

Annex 1

Sampling Plan for Studies of Title II Development Assistance Programs in Guatemala, Niger and Uganda

	Target population for Stunting	Estimated proportion of Population (A)*	Average Household size (B)*	Individuals per HH (A*B/100)	Estimated Prevalence of Stunting*	Detectable Change P2-P1	Individual Sample Size Needed	Household Sample Size Needed	Households with 10% Non-response
Guatemala	Children 0-59 months	16.0	5.2	0.8	0.48	0.06	1,694	2,695	2,965
Uganda	Children 0-59 months	19.2	5.0	1.0	0.38	0.06	1,557	2,208	2,429
Niger	Children 0-59 months	20.0	6.1	1.2	0.47	0.06	1,686	1,981	2,377

*Source for Guatemala: 1995 DHS; Niger & Uganda: 2006 DHS

Note: For Niger, nonresponse rate was assumed to be 20%

Based on the target sample sizes calculated above, ICF will sample 75 clusters with 40 households per cluster for each Title II program in Guatemala (2 programs), and 80 clusters with 30 households per cluster for each Title II program in Uganda (2 programs) and Niger (3 programs); resulting in an overall household sample size of 6,000 in Guatemala, 4,800 in Uganda, and 7,200 in Niger.

Treatment of small villages/communities on the frame

At the October 2012 meeting, two options were identified for handling communities on the sampling frame that are smaller (as defined by the number of households in the community) than the projected sample take of households per community at the second stage of sampling. These options are:

1. Eliminating such communities from the frame before sampling, provided the total of such eliminated communities constitutes a very small proportion of all households on the frame (2%-3%); or
2. Combining small communities together on the frame before sampling. It was noted that this second approach could lead to logistical issues related to travel between the combined communities (given their potential non-contiguity), should a combined pair be selected in the sample.

After assessment of the communities with less than the required number of households in each of the community lists provided by the Title II partners, it was decided to adopt the first option since these communities constituted a very small percentage (<2%) of the overall number of households for each program area.

First stage cluster sampling of villages/communities

Although surveys typically use PPS sampling (with replacement) at the first stage of sampling, the drawback of this method is that there is an inherent chance of selecting the same community twice. Therefore, an alternative method that essentially approximates PPS sampling will be used instead. For this method, communities on the frame are ordered in decreasing size (relative to the number of households within), and then separate strata are formed for large, medium, and small communities (for example). The precise number of strata that are formed depends on the

Annex 1

Sampling Plan for Studies of Title II Development Assistance Programs in Guatemala, Niger and Uganda

overall number of communities to be sampled and the number of communities on the sampling frame. Finally, an identical number of communities are randomly selected within each size stratum using systematic sampling. This method has the advantage of ensuring that large, medium, and small communities are selected in the sample. Furthermore, this method minimizes the possibility of selecting the same community twice. See Appendix C for an illustrative description of the method.

An additional consideration for the first stage of sampling is to ensure that some sampled communities fall within each of the departments and/or districts in which each program operates. In order to ensure representation in each of the specific geographic departments/districts, the “universe” of communities will first be stratified by department/district and a fraction of the total communities per Title II partner will be proportionately allocated to each department/district for sampling. Then the “universe” of communities within each department/district stratum will be ordered by decreasing household size to form “size strata”, in accordance with the method described in the paragraph above.

See Appendix D for further details of the first stage sampling methods used for each Title II program in each country.

Treatment of large villages/communities: potential segmentation of communities

ICF will work with their subcontractors to develop boundary maps for each cluster using GIS coordinates provided by the Title II partners or the Census files. Prior to the second stage sampling of households, the selected communities will either be canvassed on the ground OR Google earth maps will be produced (using GIS boundary coordinates) in order to assess the density and placement of households within the community; and to identify barriers that might prevent free access to households (such as rivers, mountains, impassable roads, etc.). After assessment of each cluster, decisions regarding segmentation of larger clusters will be made. For those clusters where segmentation is needed (i.e., in very large clusters where an enumeration would be difficult to undertake by one interviewer), interviewers will be dispersed among the segments and random starting points will be selected within each segment. Note that if segmentation is deemed necessary, sampling will take place in *all* segments. Note also that if a cluster is segmented into three parts (for example), 10 households per segment will be selected at the second stage of sampling to ensure that a total of 30 households are selected across the entire cluster as originally envisaged (for Uganda and Niger).

Second stage sampling of households

The selection of households will be done in the field using a systematic sampling method. This method entails: 1) randomly choosing a starting point between 1 and n (the sampling interval) where the household labeling 1, 2, ..., n commences at one end of the cluster; 2) conducting an interview in the first household represented by the random starting point; and 3) choosing every

Annex 1

Sampling Plan for Studies of Title II Development Assistance Programs in Guatemala, Niger and Uganda

n^{th} household from the previous one thereafter for an interview (where n is the sampling interval and equals the total number of households in the cluster divided by 30 or 40), until the entire cluster has been covered and the target number of interviews has been obtained. Specific instructions on implementing the systematic sampling will be provided to supervisors during training and in the field procedures manual.

Third stage of sampling: Multiple households within dwellings and/or polygamous households

The standard DHS definition of a household will be used: “a person or group of people who live together and share meals (“eating from the same pot”). The DHS Interviewer manual provides several definitions and examples for different types of living arrangements. For men with more than one wife (polygamous situations), the norm is to count him where he spends most of his time. So if he has three wives, but “eats from the pot” of one of the wives most often, then he would be listed as being a usual resident in her household in order to avoid duplicate counting. However, if the man is considered to be the primary farmer in all three households and the household in which he is listed is not selected for sampling (although one of the other two households is selected), then it will be preferable to interview him as the respondent for the information related to the agricultural indicators

Another common living arrangement in the Karamoja region is the so-called “big girls”. These are young women, linked to a man, who has not yet paid the dowry. These women may have children, but their work is still related to the father’s household. The “husband” is only a visitor. These women live in individual huts belonging to the father’s compound. For the Title II surveys, these women will be considered part of their father’s household as long as they are “eating from the same pot”; otherwise they will be considered as a separate, distinct household.

If there is more than one household (family) living in a dwelling, but all members of the dwelling eat from the same pot, then all members will be treated as one household and all members will be listed on the same household roster, for the purposes of sampling. However, if related households live in distinct huts in a compound dwelling (such as a manyatta in Uganda), then one household will be randomly selected from amongst them. Note that this case implies an additional stage of sampling with an associated additional sampling weight.

Fourth stage of sampling: Selection of individuals within households

The quantitative survey is broken into several modules with different individuals eligible to be interviewed, depending on the target groups relevant to the various FFP indicators. This means that, depending on the composition of a sampled household, it may or may not contain children aged 0-6 months (relevant to exclusive breastfeeding indicator), children aged 0-23 months (relevant to minimum acceptable diet indicator), children aged 0-59 months (relevant to the diarrhea, oral rehydration therapy, stunting and underweight indicators), women of reproductive age (relevant to woman’s dietary diversity and BMI indicators), farmers (relevant to agricultural

Annex 1

Sampling Plan for Studies of Title II Development Assistance Programs in Guatemala, Niger and Uganda

indicators), or heads of households/responsible adults (relevant to the household dietary diversity scale and expenditures indicators).

The household roster will be completed at the beginning of the interview, thus identifying all members of the selected household. Based on discussions during the baseline planning workshops held in each country with ICF, FANTA, FFP, and the Title II partners, the protocol for selection of individuals within households is defined as follows:

- For the modules requiring data about the household, the head of household or any responsible adult will be interviewed.
- For the children's module, the mother or caretaker of the children under age 5 will be interviewed. Data and anthropometry measures will be collected for *all* eligible children and thus *no* additional sampling weight will be needed in this case.
- For the woman's module, one woman between the ages of 15-49 will be selected. If there are multiple women eligible to be interviewed within a sampled household, a Kish grid will be used to select only one and the associated sampling weight will be computed. Note that if a pregnant or lactating woman is selected, her anthropometric measurements will be collected, although her measurement will not contribute to the estimation of the BMI indicator.
- For the agricultural module, farmers within the household who have ownership or decision-making power over all plots of land and/or livestock that are part of the "farm" will be interviewed. If in a particular household, this implies one farmer then only that farmer will be interviewed. If, however, one farmer makes decisions about the crop management practices and another farmer makes decisions about the livestock management, then each farmer will be interviewed for their respective decision-making areas. *It was also* agreed that if the primary farmer has migrated for an extended period to work outside of the household, the spouse and/or another responsible adult farmer that can answer the agricultural questions will be interviewed. For the agricultural module, since the "farm" (including the plots of land, livestock, etc) is the sampling unit and the farmer(s) are respondents in relation to the farm, there is no random selection implied and thus no additional sampling weight required.

Sampling Weights

Sample weights will be computed and used in the final data analyses. This will involve computing an overall sampling weight consisting of the product of the weights from each of the stages of sampling, as well as an adjustment to compensate for household non-response at the second and third stages of sampling. Separate sampling weights will be derived for each program area and for each target population, i.e. households, women, children and farmers.

APPENDIX A

NOTES FROM OCTOBER 10, 2012 MEETING ON SURVEY DESIGN

Summary Notes for Meeting on Survey Design for the Baseline Studies for Title II Programs in Guatemala, Uganda and Niger (written and sent by Dianna Stukel from FANTA)

Date: Wednesday, October 10, 2012, 11 am-5 pm

In Attendance: ICF International (Don Ellison, Matt Holtman, Benita O'Colmain, Suteera Nagavajara, Owen Calvert); FANTA (Megan Deitchler, Pam Velez-Vega, Diana Stukel)

Apologies: Alexandra Riboul (FFP)

Agenda Items for Discussion and Summary of Decisions Made:

1. Meaning of "Population-Based" Survey

- It was agreed that for the purposes of title II, the sample frame would include all villages/communities in which the PVOs were implementing their programs, and would exclude those in which programs were not active. From this frame, a representative sample of households and individuals (that would include both beneficiaries and non-beneficiaries alike) would be randomly drawn.

2. Choice of Indicator to drive sample size

- FANTA distributed a hand-out (see attachment) with a table that reworked some of the sample size calculations given by ICF (Table 1 in their original proposal), based on a few revised assumptions (different detectable change, different inflation factor, different household response rate). Regardless, in both the original ICF table and the reworked table, the indicator related to stunting seemed to give rise to a sample size that was both adequate and feasible (and both versions of the table gave identical sample sizes of roughly 3,000 households). Therefore, it was decided that stunting should drive the sample size calculation and that the overall sample size should be roughly 3,000 households (per PVO in each country). Given this, it could be expected that this would yield roughly 2,700 responding households (per PVO in each country), and after screening, roughly 1,700 children (per PVO in each country) under the age of 5 years old (relevant for the stunting and underweight indicators).

3. Choice of formula to drive sample size calculation

- FANTA mentioned that a somewhat different formula was used to calculate the sample size in the revised table based on a test of differences for proportions – from that which was given in the original FANTA Sampling Guide. In FANTA's opinion, the new formula is preferable to the one in the Sampling Guide because it more aptly characterizes the test of hypothesis that should be undertaken. Regardless, the original sample size formula and the new one render results that differ only negligibly (less than 5 units), and therefore, it was noted that there are no cost implications to using the new formula. Diana mentioned that she would send ICF the new formula (that would also appear in the future updated FANTA Sampling Guide) and ICF agreed to use the new formula in all future calculations.

4. Choice of inflator to determine number of households to sample to ensure the required sample size of individuals (if indicator to drive sample size is based on individual)

- FANTA noted that in their original proposal, ICF had used the sample size inflator indicated in the original FANTA Sampling Guide (1997) – but that instead they should

APPENDIX A

NOTES FROM OCTOBER 10, 2012 MEETING ON SURVEY DESIGN

use the updated sample size inflator given in the recently published Addendum (2012) to the FANTA Sampling Guide. ICF agreed with this.

5. Choice of non-response inflation factor and field strategy for non-response follow-up

- FANTA suggested that the assumed 5% non-response rate in Table 1 of the ICF proposal might be somewhat of an underestimate of what is truly required, and that 10% might be a more realistic non-response rate to expect at the household level. ICF noted that their value of 5% was based on the assumption of a short questionnaire that would not invoke much household non-response. ICF agreed to revisit this issue later.

6. Source and content for frame of clusters

- ICF noted that it would not be difficult for them to obtain Census information to build a frame of communities for Uganda and Niger. However, they noted that Guatemala would present more of a challenge. ICF asked if it might be possible to contact FTF (through FFP) to enquire after the source of the frame for FEEDBACK.
- It was noted that ICF would put together draft frames for each country, based on the broad geographic areas in which the PVOs proposed to work - and that the PVOs should then indicate in which of the communities on the draft frames they intend to implement their programs. ICF could then eliminate the communities where PVOs do not intend to implement programs from the draft frame, and use this refinement to form their final frame from which to draw communities randomly.
- ICF wondered if it would be possible to obtain from the PVOs detailed maps of the geographic areas in which they intend to work.
- ICF mentioned that they may need assistance with regards to community names on the sampling frame. They noted that, in their experience, often the same community could have more than one name. ICF hoped that the PVOs could help them arrive at a common set of names for the communities on the frame that both parties could adhere to.

7. Stratification – by PVO and potentially by other levels

- It was agreed that in each country, separate strata would be formed for each PVO, and that estimates would be produced by stratum/program as well as at the overall cross-program level within each country.
- FANTA mentioned that sometimes, PVOs implement the MCHN component of their programs in a subset of the communities where the agricultural component is implemented. Given that the intention is to spread the baseline sample across the entire geographic area where PVOs implement their programs, this could lead to results on indicators relating to MCHN showing diluted results, given that some of the sample could fall in the non-MCHN implementation zones. This was simply noted as a potential issue and but that no action need be taken other than indicating this in the analytical reporting.
- FANTA introduced the idea that there could be further stratification by other geographies (and a potential further refinement of sample allocation of communities to those strata). However, later in the meeting, it was agreed that in light of the discussion in 8 a), it might be best to put this idea to one side, as the alternative methodology discussed in 8 a) already invokes further stratification.

APPENDIX A
NOTES FROM OCTOBER 10, 2012 MEETING ON SURVEY DESIGN

8. First stage sampling

a) PPS/WR of clusters or alternative method

- FANTA introduced the issue that traditional PPS sampling *with replacement* has the disadvantage that the same communities have the potential to be selected twice in the sample. FANTA distributed a handout and discussed a possible alternative to traditional PPS WR (see attachment). For this alternative strategy, the communities on the frame are ordered in decreasing size, and then a number of separate strata are formed for large, medium and small communities. Finally, communities are randomly selected with each stratum using systematic sampling. This alternative method has the advantages of i) it being very unlikely that the same community is selected twice and ii) ensuring that some large, some medium-sized and some small communities are selected in the sample. ICF said they would review the methodology and come back at a later date with a final decision. (Note: ICF ultimately decided to adopt this methodology.)

b) Treatment of small villages/communities

- FANTA asked ICF how they intended to treat communities on the sampling frame that are smaller (as defined by the number of households in the community) than the projected sample take of households per community at the second stage of sampling. Several options were discussed including:
 - i) eliminating such communities from the frame before sampling, provided the total of such eliminated communities would only constitute a very small proportion of all communities on the frame (2%-3%); or
 - ii) combining such small communities together on the frame before sampling. It was noted that the second approach could lead to logistical issues related to travel between the combined communities (given their potential non-contiguity), should a combined pair be selected in the sample.

ICF agreed to revisit this issue after the frame was constructed.

c) Issue of segmentation of large villages/communities

- ICF mentioned that the DHS typically uses segmentation of large communities and that they also intend to do so for the Title II baseline surveys. The variant of segmentation that ICF uses divides large communities into smaller segments, and then different teams cover the divided pieces. No sub-sampling is typically undertaken.

d) Potential shadow sample of replacement village/communities

- It was mutually agreed that the discussion of this topic should be relegated to the workshops to take place in each country, given that the PVOs would be in a better position to give advice regarding communities that might be potentially problematic for interviewing – because of security, access, or other reasons.

9. Second stage sampling – systematic with listing or alternative methods

APPENDIX A

NOTES FROM OCTOBER 10, 2012 MEETING ON SURVEY DESIGN

- ICF agreed that they would adopt a second stage sampling scheme involving canvassing of the selected communities followed by the random selection of households using systematic sampling. ICF explained that they typically establish the boundaries of a selected community using GPS. They then obtain a rough count of the community by canvassing. Two teams are typically used per community, where each team consists of 5 enumerators/supervisors. ICF confirmed that alternate non-probability-based methods, such as Random Walk, would not be employed for the Title II baseline surveys.

10. Treatment of multiple households within dwellings and/or polygamous households

- It was mutually agreed that the discussion of this topic should be relegated to the workshops to take place in each country, as each country has its own specific context that needs to be considered.

11. Selection of individuals within households

- ICF agreed that they would interview all eligible individuals belonging to the target groups relevant to the various indicators for which data is to be collected within each sampled household. This means that, depending on the composition of a sampled household, it may or may not contain children aged 0-6 months (relevant to EBF indicator), children aged 0-23 months (relevant to MAD indicator), children aged 0-59 months (relevant to stunting and underweight indicators), women of reproductive age (relevant to WDDS and BMI indicators), farmers (relevant to agricultural indicators), etc. It was agreed that, when a selected household contains individuals falling in any of the target groups, all individuals in the target group relevant to the indicators for which data is to be collected will be interviewed.

12. Sampling weighting

- ICF agreed that they would compute and use sample weights in the final data analyses. This will involve computing an overall sampling weight consisting of the product of the weights from each of the stages of sampling, as well as a final adjustment to compensate for household non-response.

APPENDIX B

SUMMARY OF COMMUNITY LISTS USED FOR SAMPLING FRAMES

Guatemala

1. CRS SEGAMIL Program

- 259 communities in 2 departments, 8 municipalities
- 117 of these were new communities not on census list
- all communities have size measure (number of households); CRS provided size measures for new communities
- 33 communities with fewer than 40 households (1.9% of all households)

2. SAVE PAISANO Program

- 198 communities in 3 departments, 13 municipalities
- 28 of these were new communities not on census list
- all communities have size measure (number of households); SAVE provided size measures for new communities
- 27 communities with fewer than 40 households (1.2% of all households)

Uganda

1. Mercy Corps SUSTAIN Program

- a. 762 villages in 3 districts
- b. Size measures available for 722 villages, 548 matched to census list, 174 provided by Mercy Corps from World Food Program
- c. Size measures missing for 40 villages
- d. 61 of the 722 villages with size measures with fewer than 30 households (1% of all households with size measures)

2. ACDI/VOCA RWANU Program

- a. 402 villages in 4 districts
- b. Size measures available for 266 villages matched to census files, missing for 136 new villages
- c. 8 of the 266 matched villages with fewer than 30 households (0.6% of all households with size measures)

Niger

1. SAVE LAHIA Program

- a. 207 villages in 1 department, 5 communes
- b. 55 of these are new villages not on census list
- c. All villages have size measures (SAVE provided updated household and population counts for all 207 villages)
- d. No communities with fewer than 30 households (smallest is 39)

APPENDIX B
SUMMARY OF COMMUNITY LISTS USED FOR SAMPLING FRAMES

2. CRS PASAM TAI
 - a. Provided CRS with list of 1,824 villages in selected departments and communes, based on census files
 - b. 777 villages on census list were confirmed by CRS for program area, 422 were identified as maybes, the remaining 625 were not included in program area
 - c. CRS provided a second list with household and population counts for 897 selected villages
 - d. Of the 897 villages provided on second list, 149 of them had fewer than 30 households (0.2% of all households)
3. Mercy Corps SAWKI Program
 - a. Mercy Corps sent list of 81 villages (80 after one duplicate was removed) in 2 departments, 7 communes
 - b. 75 villages matched to census files representing 107 enumeration areas
 - c. Size measures missing for 5 villages
 - d. Of the 107 enumeration areas with household counts, 6 had fewer than 30 households (1% of all households with size measures)

APPENDIX C

ILLUSTRATIVE EXAMPLE OF MODIFIED PPS SAMPLING METHOD

Scenario: Want to select 30 villages at first stage of sampling, and 30 HH per village at second stage of sampling. Typically use PPS with Replacement (WR) at first stage and systematic sampling at second stage. Assume the frame has 60,000 HH overall

1) Alternative Method to Traditional PPS With Replacement

- Order all villages on frame in decreasing order of size (# households per village)
- Divide villages into arbitrary number of strata (say, 6), **each of roughly equal size**
- Stratum 1 has a small number of large villages and stratum 6 has a large number of small villages.
- E.g.,
 - o Stratum 1 has 10 villages each with roughly 1,000 households each (**10,000 HH overall**)
 - o Stratum 2.....
 - o Stratum 6 has 100 villages each with roughly 100 households each (**10,000 HH overall**)
- Then
 - o Stratum 1 – Select 5 villages from ordered list using systematic sampling; Select 30 HH per selected village using systematic sampling
 - o Stratum 2....
 - o Stratum 6 – Select 5 villages from ordered list using systematic sampling; Select 30 HH per selected village using systematic sampling
- What is the combined probability of selection from the combined stages?
 - o Stratum 1: $\text{Pr(overall)} = \text{Pr(stage one)} * \text{Pr(stage two)} = (5/10) * (30/1,000) = \mathbf{15/1,000}$
 - o Stratum 2:
 - o Stratum 6: $\text{Pr(overall)} = \text{Pr(stage one)} * \text{Pr(stage two)} = (5/100) * (30/100) = \mathbf{15/1,000}$
- Overall probability of selection is approximately same for each stratum: 15/1,000!!!

2) Traditional PPS WR Sampling

- No stratification, simply select 30 village with PPS WR at first stage, followed by 30 HH per selected village using systematic sampling at second stage
 - o For a village from Stratum 1,
 $\text{Pr(overall)} = \text{Pr(stage one)} * \text{Pr(stage two)} = (30 * 1,000 / 60,000) * (30 / 1,000) = \mathbf{15/1,000}$
 - o For a village from Stratum 2,....
 - o For a village from Stratum 6,
 $\text{Pr(overall)} = \text{Pr(stage one)} * \text{Pr(stage two)} = (30 * 100 / 60,000) * (30 / 100) = \mathbf{15/1,000}$

Overall Advantages of Alternative Method

- Approximately same overall probabilities using alternative method as with PPS WR followed by systematic sampling
- Closer to PPS without replacement sampling (PPS WOR) in that much less likely to select the same village twice (Key advantage)
- Very easy to implement since essentially systematic sampling at first stage

APPENDIX D
SUMMARY OF FIRST-STAGE SAMPLING METHODS FOR EACH TITLE II PROGRAM

Guatemala

Aim: Select 75 communities for each Title II program (SEGAMIL and PAISANO)

1. Remove communities with fewer than 40 households from the sampling frame.
2. Proportionately allocate the 75 communities to be sampled for each program to the departments where each program operates (two departments for CRS/SEGAMIL; three departments for SAVE/PAISANO); use total number of households in each department to determine the allocation.
3. Order the communities within each department/stratum by decreasing household size.
4. Examine the distribution of number of households for the communities within each department/stratum to determine appropriate cut-off points for the “size” strata to be used at the first stage of sampling (stratified systematic sampling)

Uganda

Aim: Select 80 villages for each Title II program (SUSTAIN and RWANU)

1. Remove villages with fewer than 30 households from the sampling frame.
2. Proportionately allocate the 80 villages to be sampled for each program to the districts where each program operates (three districts for Mercy Corps/SUSTAIN; four districts for ACDI/VOCA/RWANU); use total number of households in each district to determine the allocation.
3. For villages with size measures, select villages using the same sampling procedure described for Guatemala above in steps 3 and 4.
4. For villages without size measures, use stratified (by district) systematic sampling.
5. To determine the number of villages (from amongst the total of 80 for each program) to sample for each group (those with size measures and those without) at steps 3 and 4, calculate the proportion of total villages for each group and then multiply this number times 80. For example, 177 villages out of 207 have size measures for the SUSTAIN program which represents 86% of the 207 villages. So, 68 villages (0.86 times 80) will be sampled from this group, and the remaining 12 villages will be sampled from the group without size measures.
6. Sampled villages with unknown numbers of households may or may not meet the criteria for 30 or more households. Village sizes for these villages will be determined in the field. Any sampled villages found to have fewer than 30 households (after verification from the field), will be replaced with villages from the same group of villages (those without size measures). The sampling weight will be adjusted to remove these villages since they would not have been included had the number of households been known at the time of sampling.

Niger

APPENDIX D
SUMMARY OF FIRST-STAGE SAMPLING METHODS FOR EACH TITLE II PROGRAM

Aim: Select 80 villages for each Title II program (LAHIA, PASAM-TAI and SAWKI)

- Remove villages with fewer than 30 households from the sampling frame.
- For CRS and SAVE programs, use the same sampling method described for Guatemala above in steps 2 and 3 (and noting that there is 1 department for SAVE and 2 for CRS).
- For Mercy Corps, select all villages since there are only 80 villages on the village list provided by Mercy Corps and 80 are required to be sampled. Of the 80 villages on the list provided by Mercy Corps, 75 are represented by 102 enumeration areas on the Niger census files and 5 villages did not match to the census file. To meet the criteria for selecting the 80 enumeration areas to be surveyed, the following selections are made:
 - a. 9 villages are represented by 2 enumeration areas each. One enumeration area is randomly selected for each of these 9 villages, giving a total of 9 enumeration areas sampled out of 18. First stage sampling probability is 0.50.
 - b. 3 villages are represented by 3 enumeration areas each. One enumeration area is randomly selected for each of these 3 villages, giving a total of 3 enumeration areas sampled out of 9. First stage sampling probability is 0.33.
 - c. 2 villages are represented by 5 enumeration areas each. Two enumeration areas are randomly selected for each of these two villages, giving a total of 4 enumeration areas sampled out of 10. First stage sampling probability is 0.40.
 - d. Of the 65 remaining enumerations areas representing 65 villages, 59 villages with 30 or more households are selected (6 villages with less than 30 households are not sampled). First stage sampling probability is 1.0.
 - e. All 5 villages that did not match to census files are selected. First stage sampling probability is 1.0.