

(Updated by Natalie 8/22/16)

## Analysis Plan for Selecting Target Households

1. There were **33,469** observations after appending the 5 separate data sets together
2. 8 observations were dropped because there were no values for any of the variables bringing the total amount of households down to **33,461**
  - a. 477 households (as defined by 'householdid') had duplicates - each of the 477 households had at least one duplicate but for 27 households there were two duplicates
  - b. 914 households (not removing duplicate households) were missing or had incorrect data inputted for at least one of the three criteria - cultivable land, missed meals in the last lean season, or migration in the last three years for the first working member
  - c. 582 households (not removing duplicate households) had an issue whereby they did not provide the prior migration history of at least one of the working members listed in the household
3. 1,381 households (not removing duplicates) had at least one of the three eligibility criteria missing/inputted incorrectly and/or had an issue where the migration history for the listed working members were not provided - these households were removed from the eligibility tests and **32,080** households (not removing duplicates) were considered for eligibility
  - a. 10,584 households (about 33% of the 32,080 households) were eligible for the program
  - b. The 1,381 households missing eligibility information/migration issue were assigned a value of “.”
4. Removal of duplicate households was addressed after eligibility was determined: For the households with duplicate observations, the eligibility for each observation within one household was checked for consistency - if one household had inconsistent results for eligibility (e.g. one observation states 'eligible' and the other states 'ineligible' or one observation states 'eligible' and the other is 'missing'), the household was marked as 'eligible'
  - a. The process for dealing with duplicates is provided in more detail in the subsection below, as well as a suggestion for how we move forward with these households
5. After applying all of the steps above, **10,427** households were eligible (about 33% of households which provided information for the eligibility criteria), 21,201 households were ineligible and 1,347 households had missing information for the eligible criteria - in total there were **32,975** households in our census
  - a. Villages with targeting: No additional steps were taken - **7,918** households were eligible, 15,827 households were ineligible and 880 households had missing information for the eligible criteria (total households: 24,625)

- b. Villages without targeting: The process for the selection of villages to be included in this group is provided in detail below - in total 15 villages were selected with a total of **8,350** households

A list of eligible households can be found here: [link redacted]

## How duplicate households were handled

1. 477 duplicate households were tagged in a variable called 'dupl'
2. Once eligibility was assigned to each observations (as described above), I created a variable ('eg') that added up the numeric values assigned to eligibility for each 'householdid' - eligible observations were assigned a value of '1' and ineligible households were assigned a value of '0'
  - a. For example - if a householdid had a duplicate and one observation for that householdid showed the household as eligible and another observation for that householdid showed the household as ineligible, the total for that household would be '1' and a value of '1' would be assigned to 'eg' for both observations in that householdid
3. I created a second variable, 'incon', to tag if there were inconsistent results for eligibility within a household
  - a. Households were assigned a value of '1' for this variable (indicating there was inconsistent eligibility) if 'eg' = 1 and the household had one duplicate, if there were two duplicates for one household (3 observations in total) a value of '1' was assigned if 'eg' = 1 or 2 and the household had 2 duplicates
4. I then created a new eligible variable that applied the eligibility criteria but if the duplicate household had inconsistent eligibility then it was decided that if a duplicate household was found eligible in at least one of the observations (i.e. 'incon' = 1) the household was considered eligible
  - a. Households for which all duplicate observations had the same eligibility were not affected by this change
5. I then created a tag variable that tagged one observation for each household ID - this tag was used when selecting eligible households (i.e. "eligible if tag = 1")
  - a. For households without duplicates, the only observation per household was tagged with a value of '1'
  - b. For households with duplicates, one observation (out of 2 or 3 per household) was tagged with a value of '1' and the other observations had a value of '0'

\*\*\*Note: For some households, duplicate observation(s) under one household ID had the same information provided for household working members and eligibility criteria. For other duplicate households, they information varied substantially across duplicate observations. For the purposes of targeting, we did not sort through these differences and suggest MOs and/or BMs check household information to ensure we have the correct data.

## How missing values for eligibility criteria were handled

- A variable was created to indicate whether a household had one of the following issues:
  - Missing values for cultivable land: 592 households
  - Missing values for the meal variable: 523 households
  - Missing values for the migration variable for the first working member: 551 households
  - Incorrectly entered data (e.g. the value to whether the second working member had ever migrated was '9' when the only two values acceptable were '0' and '1')
  - If a household listed a working member but did not provide migration information for previous years
    - For some households zero was entered for a working member's name - all "0" were replaced to 'missing' to indicate that there was no second, third, fourth, etc. working member
- This variable was used when applying eligibility criteria

## Sampling of villages to be included in the sample without targeting

1. 60 villages (in 12 branches) were provided for the 'no targeting' approach, of which only 15 villages could be selected
  - a. It was noted that RDRS did not want us to administratively overwhelm the branches, therefore villages with more than 500 households were removed from the list of 82 villages for a total of 60
2. Correlation tests between village eligibility rates, distance/time from branch and number of households per village were carried out
  - a. No. of households and eligibility rates per village were both negatively correlated with the distance/time from branch, with correlations of -0.3450 and -0.4757 respectively
  - b. Correlation between no. of households and eligibility were weakly positively correlated, with a correlation of 0.1574
3. It was then decided to move forward with stratifying the available 60 villages on branch and eligibility rates per village
4. The distribution of eligibility rates in the subsample of 60 villages revealed the median eligibility rate was 31.6% - the median was used to create a variable which classified villages as having either a 'low' eligibility rate (below 31.6) or a 'high' one (greater than or equal to 31.6)

		Low	High	Total
1	BDS	1	3	4
2	BOS	1	6	7
3	ESH	11	2	13
4	GAN	2	1	3
5	KAL	3	0	3
6	KHU	0	1	1

7	NAK	0	2	2
8	NAS	2	4	6
9	PAT	0	7	7
10	RAS	9	0	9
11	SON	1	1	2
12	TAR	0	3	3
	Total	30	30	60

5. Using this discrete cut-off, we randomly selected two villages from each branch - one village from the 'low' category and one village from the 'high' category' - this resulted in 23 villages selected (10 in the 'low' category and 13 in the 'high' category)
  - a. If the branch only had one village in a category, that one village was selected
  - b. If the branch had no villages in a certain category, then two villages were randomly selected from the opposite category (e.g. in KAL - two villages were randomly selected from the 'low' category)
    - i. For KHU - there was only one village to select therefore KHU is the only branch with one village
6. The remaining 7 villages were randomly selected from the remaining villages in either the 'low' or 'high' category
  - a. 5 villages were randomly selected from the 20 remaining villages in the 'low' category
  - b. 2 villages were randomly selected from the 17 remaining villages in the 'high' category

## Census Data

Census data files and aggregated data files can be found here: [link redacted]

## Do Files for Targeting

Do files for targeting eligibility and village selection for 'no targeting' can be found here: [link redacted]