STEP Survey Weighting Procedures Summary
(Based on The World Bank Weight Requirement)

China-Yunnan

October 11, 2013

Table of Contents

1   Survey Design Overview

The China-Yunnan survey firm implemented a partial literacy assessment design. The partial assessment required each selected person to attempt to complete a General Booklet comprising Reading Components and a set of Core Literacy Items. The partial assessment sampling objective was to have a minimum of about 2000 selected persons attempt the General Booklet.

The target population for the China-Yunnan STEP survey comprised all non-institutionalized persons 15 to 64 years of age (inclusive) living in private dwellings in urban areas of **Kunming** at the time of data collection.

The sample frame for the selection of first stage sample units was the Excel file 'sampling frame for STEP _CHINA' that was provided by the China-Yunnan survey firm. The frame is a complete list of first stage sampling units in the urban areas of Kunming. The source of this sample frame is the National Population Census, November, 2010.

The sample frame includes 5564 PSUs in 299 Census Enumeration Areas. According to the sample frame, there are 1,067,256 households in the 5564 PSUs.

The China-Yunnan sample design was a 3 stage cluster sample design.

First Stage Sample
The primary sample unit (PSU) is a Census Enumeration Area (CEA) Block. The sampling objective was to conduct interviews in 135 CEA Blocks. At the first stage of sample selection, 27 additional PSUs were also selected as reserve PSUs to be used in the event that it was impossible to obtain any interviews in one or more of the initial PSUs. A total of 162 PSUs were selected with probability proportional to size, where the measure of size was the number of households in a PSU. Subsequently, from the file of 162 sampled PSUs, a PPS sample of 135 PSUs was selected to be the 'Initial' PSU sample. Note that none of the 27 reserve PSUs was activated during data collection.

Second Stage Sample
 The second stage sample unit (SSU) is a household. The sampling objective was to obtain interviews at 15 households within each selected PSU. At the second stage of sample selection, 30 households were selected in each PSU using a systematic random method. The 30 households were randomly divided into 15 'Initial' households, and 15 'Reserve' households that were ranked according to the random sample selection order.

Third Stage Sample
The third stage sample unit was an individual aged 15-64 (inclusive). The sampling objective was to select one individual with equal probability from each selected household.

2   Data Collection Outcomes

The final survey weights depend on the outcome from the attempt to obtain an interview at each selected household. The treatment of each sampled case in the weighting process depends on the result code assigned to the sampled household and the result code assigned to the selected person. The result codes used in STEP are listed in Table 1. A variable called '**FinResultCode**' with these code values was created during the weighting process and a '**FinResultCode**' was assigned to each sampled case.

Table 1: *Partial Literacy Assessment* Result Codes for Sampled Households and Selected Persons	
I. Household-level Codes	
Code	**Code Description**
01	Household Refusal
02	Household unusual circumstance (e.g., death in family, illness, fire in dwelling, etc.)
03	No knowledgeable household member found after 3 visits (only a child, non-competent adult, etc.)
04	Temporarily absent/unavailable for field period (information from other persons)
05	No competent household member to interview (because of severe illness, mental disability, etc.)
06	Language problem – no household member spoke a language understandable by survey team, no translator available
07	Dwelling could not be found / selected address has no household
08	Vacant dwelling
09	Dwelling not habitable/dwelling destroyed/dwelling converted to commercial use
21	Household Refusal to continue Household Module (i.e., begun but not complete)
22	Household unusual circumstance (e.g., death in family, illness, fire in dwelling, etc.) prevented continuance
31	No household member in the eligible range of 15-64
94	Case dropped due to data quality issue
95	Sample case justifiably not activated, e.g., reserve sample household not needed , sample goal achieved
96	General non-response; Reason unspecified by survey firm
97	Household not interviewed due to sample selection error – wrong household selected in its place
II. Person-level Codes	
Code	**Code Description**
	Individual modules 2-7 not begun or unacceptable interview
32	Refusal by Selected Person to begin modules 2-7
33	Unusual circumstance - Selected Person (e.g., death in family, illness, fire in dwelling, etc.)
34	Refusal - Another household member refused to let selected individual do individual modules
35	Unable to contact Selected Person after three visits during field period
36	Temporarily absence/unavailability of Selected Person for field period (info from household member)
37	Hearing Problem - Selected Person is deaf or so hard-of-hearing, no translator available
38	Learning or mental disability (includes emotional conditions like severe depression) of Selected Person
39	Speech impairment of Selected Person prevented participation, no translator available
40	Language problem – Selected Person's language not understood by survey team, no translator available
98	Sample selection error – an incorrect eligible person was selected
99	Sample selection error – an ineligible person was selected
	Individual modules 2-7 begun but not completed
51	Refusal by Selected Person to continue modules 2-7
52	Unusual circumstance - Selected Person
	General Booklet (module 9) not begun
61	Unable to read or write in language of General Booklet – Selected Person refused to begin
62	Refusal by Selected Person to begin General Booklet
63	Unusual circumstance - Selected Person
64	Blindness or visual impairment of Selected Person prevented completion of General Booklet
65	Physical disability of Selected Person prevented completion of General Booklet
	General booklet (module 9) looked through or attempted but not completed
71	Selected Person looked at Booklet but unable to read & write the language of Booklet - did not begin
72	Refusal by Selected Person to continue the General Booklet
73	Interruption – unacceptably long, General Booklet administration terminated
74	Unusual circumstance - Selected Person
75	Attempted General Booklet but no answers marked

Table 1: *Partial Literacy Assessment* Result Codes for Sampled Households and Selected Persons	
76	Marked some items but did not attempt all items in General Booklet
77	Marked some items & attempted all items in General Booklet
78	Marked all items in General Booklet

3   Weighting Overview

While the China-Yunnan three-stage stratified cluster design greatly enhanced the operational feasibility of data collection, it resulted in differential probabilities of selection for the selected persons. Consequently, each selected person in the survey does not necessarily represent the same number of persons in the target population. To account for differential probabilities of selection due to the nature of the design and to ensure accurate survey estimates, STEP requires a sampling weight for each person that participated in the survey.

In general, the objectives of the STEP weighting are to construct a set of survey weights to,
1) compensate for unequal probabilities of selection;
2) compensate for household-level non-response and person-level non-response;
3) adjust the weighted sample distribution for key variables of interest (for example, age, gender, education) so that it conforms to a known population distribution for these variables.

The general weighting procedure for the China-Yunnan STEP survey required the following tasks.

1) Creation of a data file to input into the weighting process;

2) **World Bank (WB) Weight Requirement**:

   Create survey weights for sampled cases of households and persons that provided sufficient data to be considered a participant in the survey. This requirement does not necessarily include the completion of an assessment General Booklet, nor does it necessarily include the completion of all household and individual questionnaire modules.
   a) Calculation of a PSU weight for 135 sampled PSUs;
   b) Calculation of a household weight for each sampled household;
      i) Calculation of a household-level non-response adjustment independently for each PSU.
   c) Calculation of a person weight for each selected person (SP);
      i) Calculation of a non-response adjustment independently for each sampled person.

3) The required output from the weighting process is a final China-Yunnan data file with the survey design weights (i.e., for each sampled PSU, household, person) appended to each data record.

4 Files - Weighting Process

4.1 Input Files

The key files for the weighting process are the original sample file created by the WB survey methodologist and the data files provided by the China-Yunnan survey firm. The objective was to use these key files to create a master data file for input into the weighting process.

4.1.1 Sample File

The SPSS file 'CHN01-Final Sample (01-06-2012)-Initial PSUs-HHs I and R-with revised PSUIDnew' is the file of initial sample PSUs. This file contains the sample variables outlined in Table 2.

Table 2: Record Layout for the SPSS file 'CHN01-Final Sample (01-06-2012)-Initial PSUs-HHs I and R-with revised PSUIDnew'			
Variable Name	**Type of Variable**	**Width**	**Variable Description**
Country	String	3	Country Code=CHN
SMPLID	Restricted Numeric	4	Case #, 0001 to 4050
CEAID	Numeric	12	CEA Identification #, original Sample frame variable
BlockID	String	2	Block Identification #, original Sample frame variable
PSUID	String	6	PSU Identification #, original Sample frame variable
Householdnumbers	Numeric	5	Number of PSU households, original Sample frame variable
PSUIDnew	Restricted Numeric	3	Sampled PSU Identification #, 101 to 235
HHsmpSeq	Restricted Numeric	2	Household Sample Sequence #, corresponds to the field systematic sample sequence numbers
HHsmpType	Numeric	1	Household Sample Indicator 1 = Initial Sample 2 = Reserve Sample
HHsmpOrder	Restricted Numeric	2	Household Sample Order 00 = Initial Sample, no household order necessary '01-15' = Reserve Sample household order

4.1.2 Data Files from Survey Firm & World Bank

The survey firm provided the China-Yunnan STEP data files to the World Bank core team. The World Bank team reviewed the data files and revised data variables to be consistent with the data editing requirements for the weighting process.

The following data files were provided by the World Bank core team:

1) The Stata data file 'ch_rst_code.dta' was provided by the China-Yunnan survey firm as a clean file of STEP data.
   a) This file contains 2025 sample cases.
   b) This file is a primary dataset for input to the weighting process.

2) The Excel file 'STEP - Yunnan - Replaced Households-for wtg-OP1' is the file of non-responding households for which a reserve household was activated.
   a) This file was created from the Excel file 'STEP - Yunnan - Replaced Households_Reviewed_TR_2', a file of non-responding households originally provided by the survey firm; the information that was not needed for the weighting process was stripped from this file to create 'STEP - Yunnan - Replaced Households-for wtg-OP1'.
   b) This file contains 21 sample cases.

3) The Excel file 'Yunnan_Stage 2_Sampling Unit' provides the total number of listed households for each of the 135 selected PSUs from the 2010 Census household listings.

4.1.3 File Preparations

The following steps were undertaken to prepare the China-Yunnan data files for the weighting process.

Step1: File Conversions

Since the weighting process is carried out using SPSS software, the Stata data file 'ch_rst_code.dta' and the Excel files 'STEP - Yunnan - Replaced Households-for wtg-OP1' and 'Yunnan_Stage 2_Sampling Unit' were converted to SPSS data files. Hereafter, reference to the following file names will imply the SPSS version of these files.

1) 'ch_rst_code'
2) 'STEP - Yunnan - Replaced Households-for wtg-OP1'
3) 'Yunnan_Stage 2_Sampling Unit'

Table 3 provides the frequency counts by 'Result Code of Interview' for the China data file 'ch_rst_code'.

Table 3: Result code of interview- China-Yunnan Data File 'ch_rst_code.dta'

Result Code & Description	Frequency	Percent
52-Interview termination-SP unusual circumstance	1	.0
62-SP refusal to begin GB	19	.9
64-SP blindness or visual impairment-unable to do GB	1	.0
72-SP refused to continue GB after starting	4	.2
73-GB interruption	2	.1
74-GB terminated due to unusual circumstance	4	.2
76-SP marked some GB items-did not attempt all items	30	1.5
77-SP marked some GB items & attempted all GB items	71	3.5
78-SP marked all GB items	1893	93.5
Total	**2025**	**100.0**

Step2: Create China-Yunnan Data File of Finalized Cases

1) Match 1: Add non-responding cases to 'ch_rst_code'.
   a) Match 'ch_rst_code' and 'STEP - Yunnan - Replaced Households-for wtg-OP1'

      i) 21 non-participating cases were added to the file 'ch_rst_code'
      ii) Revised the value labels for the variable 'm11_q07', the Result of Interview Code, to correspond to the Codes and code descriptions in Table 1 of this document.
         ◆ **SPSS Output File**: 'ch_rst_code-1'

      iii) Created the variable 'FinResultCode', Final Result Code, and defined its value labels to correspond to the Codes and code descriptions in Table 1 of this document.

      iv) Revised Final Result Code, i.e., variable 'FinResultCode', as follows:
         ◆ IF (id_ques='CH12210') FinResultCode=99.
         ◆ IF (id_ques='CH21227') FinResultCode=98.
         ◆ IF (id_ques='CH16725') FinResultCode=98.
         ◆ IF (id_ques='CH16610') FinResultCode=97.
         ◆ IF (id_ques='CH16714') FinResultCode=97.
         ◆ IF (id_ques='CH16717') FinResultCode=97.
         ◆ IF (id_ques='CH18925') FinResultCode=97.
         ◆ IF (id_ques='CH20730') FinResultCode=97.

      v) **SPSS Output File**: 'W01-CHN Final Cases(07-NOV-2012)'

Step 3: Create the China-Yunnan data file of finalized cases from the survey firm.
2) Match 2: Add Household Counts from Household Listing Operation
   a) Match 'W01-CHN Final Cases(07-NOV-2012)' and 'Yunnan_Stage 2_Sampling Unit' to add the variable indicating the number of listed households per selected PSU cases to W01-CHN Final Cases(07-NOV-2012)'.

      i) Variable added – 'total#hhldslisted'

      ii) **SPSS Output File**: 'W01-CHN Final Cases(07-NOV-2012)', the file of finalized cases.
         ◆ This file includes all 2046 finalized cases from the survey firm. Each case includes a final result code. Table 4 provides the breakdown of the 2046 cases by Final Result Code.

Table 4: China Step Data - Number of Finalized Cases by Final Result Code		
Final Result Code & Description	**Frequency**	**Percent**
03-No knowledgeable household member found after 3 visits	4	.2
04-Temporarily absent/unavailable for field period	1	.0
08-Vacant dwelling	2	.1
09-Dwelling not habitable/dwelling destroyed/dwelling converted to commercial use	3	.1
31-No eligible household member aged 15-64	3	.1
52-Interview termination-SP unusual circumstance	1	.0
62-SP refusal to begin GB	19	.9
64-SP blindness or visual impairment-unable to do GB	1	.0
72-SP refused to continue GB after starting	4	.2
73-GB interruption	2	.1
74-GB terminated due to unusual circumstance	4	.2
76-SP marked some GB items-did not attempt all items	30	1.5
77-SP marked some GB items & attempted all GB items	69	3.4
78-SP marked all GB items	1887	92.2
96-General non-response; Reason unspecified by survey firm	8	.4
97-Sample selection error – wrong household selected in the field	5	.2
98-Sample selection error – incorrect eligble person selected by the interviewer	2	.1
99-Sample selection error – ineligble person selected by the interviewer	1	.0
Total	**2046**	**100.0**

Step4: Create the final data file of all sampled cases with a final result code for each case.
3) Match 3: Add variables from the original sample to the China data
   a) Match 'CHN01-Final Sample (01-06-2012)-Initial PSUs-HHs I and R-with revised PSUIDnew' and 'W01-CHN Final Cases(07-NOV-2012)' to create the file of finalized cases.

   b) **SPSS Output File**: 'W02-CHN Final Sampled Cases(07-Nov-2012)'
      i) This output file contains 4050 sample cases with a final result code for each case.
      ii) This file is the primary file to for the process of creating the survey weights.
         ◆ Table 5 provides the list of key variables for weighting that are included in 'W02-CHN Final Sampled Cases(07-Nov-2012)'
         ◆ Table 6 provides the breakdown of the 4050 cases by Final Result Code and by Household Sample Type.
            ▪ **Assumption: For weighting purposes, it is assumed that the 2004 reserve sample cases with Final Result Code '**95-Sample case not activated**' on the original sample file are cases that were never activated during the data collection operation.**

The survey firm provided a total of 2046 finalized cases that include 2025 'initial' sample households and 21 'reserve' households that were activated for non-responding sampled households.

For weighting purposes, a finalized case is a sample household for which a final result code has been assigned. Ideally, there should be a final result code for each of the 4050 sampled households on the original sample file. During the file preparations for weighting, a final result code was assigned to the 4050 sampled cases.

Table 5: Key Variables in Primary Input File for Weighting Process		
Variable Name	**Source**	**Description**
Country	Sample File	Country abbreviation code
SMPLID	Sample File	Sampled case ID#
CEAID	Sample File	CEA ID
BlockID	Sample File	Block ID
householdnumbers	Sample File	Sample Frame # of households in PSU
PSUIDnew	Sample File	PSU ID - Sampled PSUs
HHsmpSeq	Sample File	Household Sample Sequence #s 01 to 30
HHsmpType	Sample File	HH Sample Indicator
HHsmpOrder	Sample File	Sample activation order
id_ques	China File	Questionnaire Identifier
cluster	China File	Household Id
hhn	China File	Household Number
check_digit	China File	check digit
m1c_a	China File	NUMBER A-# of eligible persons
m1c_b	China File	NUMBER B- listing # of selected person
m11_q07	China File	Result code of interview
FinResultCode	Created during file preparations	Final Result Code
total#hhldslisted	China File	total # of listed households in PSU

Table 6: China Step Data - Number of Finalized Cases by Final response Code & HH Sample Type			
Final Result Code & Description	**Initial Sample**	**Reserve Sample**	**Total**
03-No knowledgeable household member found after 3 visits	4	0	4
04-Temporarily absent/unavailable for field period	1	0	1
08-Vacant dwelling	2	0	2
09-Dwelling not habitable/dwelling destroyed/dwelling converted to commercial use	3	0	3
31-No eligible household member aged 15-64	3	0	3
52-Interview termination-SP unusual circumstance	1	0	1
62-SP refusal to begin GB	19	0	19
64-SP blindness or visual impairment-unable to do GB	1	0	1
72-SP refused to continue GB after starting	4	0	4
73-GB interruption	2	0	2
74-GB terminated due to unusual circumstance	4	0	4
76-SP marked some GB items-did not attempt all items	29	1	30
77-SP marked some GB items & attempted all GB items	69	0	69
78-SP marked all GB items	1875	12	1887
95-Sample case not activated	0	2004	2004
96-General non-response; Reason unspecified by survey firm	5	3	8
97- Sample selection error – wrong household selected in the field	0	5	5
98-Sample selection error – incorrect eligible person selected by the interviewer	2	0	2
99-Sample selection error – ineligible person selected by the interviewer	1	0	1
Total	**2025**	**2025**	**4050**

4.2 Final Output Files from Weighting Process

The output files from the weighting process are,

1) W03-CHN-STEP HH Survey Weighted Cases (07-Nov-2012)
   a) This file contains the applicable PSU weights, household weights, person weights and other associated variables that were created during the weighting process.
   b) The file contains,
      i) 2020 participating households with a final household weight
         ◆ For participating households, the weighted sum of the household weights is 1,063,933
      ii) 2017 participating selected persons with a final person weight
         ◆ For participating individuals, the weighted sum of the person weights is 2,612,811

2) CHN-W01-STEP(HH) Weighted Data-Basic Weights (07-Nov-2012)
   a) This file contains the original data variables from the China-Yunnan data file 'ch_rst_code' and the applicable PSU weights, household weights, person weights and other associated variables that were created during the weighting process.
   b) The file contains a data record for each of the 2046 cases provided by the survey firm.
   c) In addition to the China-Yunnan data file original variables, the weighted data file includes several additional variables that were created during the weighting process. The essential variables that are needed for estimation of survey results are included in the final weighted data file.
      i) Table 7 provides the list of variables from the weighting process that are included in the final weighted data file.
      ii) Table 8 provides the list of variables from the weighting process that are not included in the final weighted data file.
      iii) Table 9 provides the frequency count of result codes for participating households in the final weighted data file 'CHN-W01-STEP(HH) Weighted Data-Basic Weights (07-Nov-2012)'.

Table 7: Variables Created or Used During Weighting Process - _Added to China Data File_

Variable Name	Description	Values
HHsmpType	HH Sample Indicator	n/a
HHsmpSeq	Household Sample Sequence #s for field	
HHsmpOrder	Reserve sample Activation Order	00 = Initial Sample, no order necessary '01-15' = Reserve Sample activation order
householdnumbers	Sample Frame # of households in PSU	n/a
total#hhldslisted	total # of listed households in PSU	n/a
FinResultCode	Final Result Code	Refer to Table 1 for code descriptions
FinRespCat_HH	Final Response Category-Households	1-Participant HH 2-Ineligible HH 3-Non-participant Eligible HH 4-Non-participant Unknown Eligibility HH
FinRespCat_SP	Final Response Category-Selected Person	1-Participant SP 2-Non-participant, Eligible 3-Non-participant, Ineligible 9-Not applicable - non-participating HH
W1_PSUwt	Basic PSU Weight	n/a
W_FinHHwt	Final Household Weight Use for estimation of Household characteristics	n/a
W_FinSPwt	Final Person Weight Use for estimation of Person characteristics	n/a

Table 8: Variables Created or Used During Weighting Process – Not *Added to China Data File*

Variable Name	Description	Values
PSU_FIN#	Count of 'FinRespCat_HH' in PSU	n/a
HH_k	# of participant HHs in PSU	n/a
HH_x	# of Ineligible HHs in PSU	n/a
HH_z	# of Non-participant Eligible HHs in PSU	n/a
HH_u	# of Non-participant Unknown Eligible HHs in PSU	n/a
W2_HH	Household Weight within PSU	n/a
W2_HHwt_basic	Basic Household Weight	n/a
Elig_Prop	Proportion of Eligible HHs - known eligibility	n/a
R_adj_HH	Response adjustment ratio for HHs	n/a
EligPersCnt	Eligible Person Count in Sampled HH	n/a
W3_SP	Selected Person n weight within household	n/a
R_adj_SP	Response adjustment for selected person	n/a
PSUIDnew	PSU ID - Sampled PSUs	n/a
PSU_EPsum	Total Eligible Persons-PSU Participant Households	n/a
PSU_PPcumul	Total Eligible Persons-PSU Participant SP Households	n/a

Table 9: China Final Weighted Data File – Frequency of Final Result Codes FOR CASES WITH FINRESPCAT_HH=1, i.e., participating households

	Frequency	Percent
52-Interview termination-SP unusual circumstance	1	.0
62-SP refusal to begin GB	19	.9
64-SP blindness or visual impairment-unable to do GB	1	.0
72-SP refused to continue GB after starting	4	.2
73-GB interruption	2	.1
74-GB terminated due to unusual circumstance	4	.2
76-SP marked some GB items-did not attempt all items	30	1.5
77-SP marked some GB items & attempted all GB items	69	3.4
78-SP marked all GB items	1887	93.4
98-Sample selection error – incorrect eligible person selected by the interviewer	2	.1
99-Sample selection error – ineligible person selected by the interviewer	1	.0
Total	2020	100.0

5   Weight Calculations based on World Bank (WB) Weight Requirement

The input file for the process of calculating the weights based on the World Bank (WB) weight requirement is 'W02-CHN Final Sampled Cases(07-Nov-2012)'.

5.1   First-stage Weight (PSU Weight)

The basic first stage weight, i.e., the PSU weight, is based on a sample size of 135 PSUs. Although 162 PSUs were initially selected, only the 135 'initial' PSU sample was actually activated during data collection; none of the 27 'reserve' PSUs was used for the China STEP data collection. Therefore, the STEP survey weight is based on the probability of selection for each of the 135 'initial' PSU sample.

For the China-Yunnan PPS sample, the probability of selection of the jth PSU, say p_j , may be represented as,

    1) $p_j = \frac{y * m_j}{M}$

where,

    y - the number of sampled PSUs (i.e., y=135 CEAs in the case of the China-Yunnan STEP Survey),

    m_j -the measure of size for the jth PSU, in the case of the China-Yunnan STEP Survey the measure of size for the first stage selection of PSUs was the sample frame number of households (i.e., weighted data file variable '**householdnumbers**') in each PSU in the target population,

    M - the total number of households in the target population, is represented by

    2) $M = \sum_{j=1}^{Y} m_j$

where,

    Y - the total number of PSUs in the target population; U=5564 for the China-Yunnan survey.

    M=1,067,256 is the sample frame count of households for the China-Yunnan STEP survey.

The basic first stage weight for the jth PSU, W_{1j} (i.e., weighted data file variable '**W1_PSUwt**'), is calculated as the inverse of its probability of selection:

    3) $W_{1j} = \frac{1}{p_j} = \frac{M}{y * m_j}$

    ➢ **The basic first stage PSU weight is appended to the STEP data file.**

Example 1: First Stage Weight

The basic first stage weight for the China-Yunnan PSU # 126 which contains 200 sample frame households is calculated as follows:

$$W_{1j} = \frac{1}{p_j} = \frac{M}{y * m_j} = \frac{1,067,256}{135 * 180} \approx 43.92$$

The basic first stage weight is similarly calculated for each of the 135 PSUs.

5.2 Second-stage Weight (Household Weight)

The basic second stage weight for the kth household in PSU 'j', W_{2jk}, is calculated as follows:

$$4) \qquad W_{2jk} = \frac{m'_j}{n_j}$$

where,

m'_j - the number of **_listed_** households (i.e., weighted data file variable '**total#hhldslisted**') in the jth PSU,

n_j - the number of households sampled in the jth PSU, (i.e., n_j =30 households in each China-Yunnan sampled PSU).

Example 2: Second Stage Basic Weight

The basic second stage weight for each of the 30 sampled households in the China-Yunnan PSU # 126 which contains 200 **_listed_** households is calculated as follows:

$$W_{2jk} = \frac{m'_j}{n_j} = \frac{200}{30} \approx 6.6667$$

The basic second stage weight is similarly calculated for each of the 135 PSUs.

5.2.1 Second-stage Response Adjustment

For treatment in the weighting process the data collection outcomes for the sampled households may be summarized in four possible categories: participant, ineligible, non-participant with known eligibility, non-participant with unknown eligibility.

Participant - A participating household '**k**' is an <u>eligible</u> sampled household that provides sufficient data to be considered a participant in the survey. The sampled households '**k**' with person-level result codes 32-40, 51-78 and 98-99 satisfy the requirement for classification of the household as a survey participant. The result codes and corresponding code descriptions for participating households '**k**' are as follows:

Code	Code Description
I.	**Household-level Codes – Participant, Eligible Household**
	Individual modules 2-7 <u>not</u> begun
32	Refusal by Selected Person to begin modules 2-7
33	Unusual circumstance - Selected Person (e.g., death in family, illness, fire in dwelling, etc.)
34	Refusal - Another household member refused to let selected individual do individual modules
35	Unable to contact Selected Person after three visits during field period
36	Temporarily absence/unavailability of Selected Person for field period (info from household member)
37	Hearing Problem - Selected Person is deaf or so hard-of-hearing, no translator available
38	Learning or mental disability (includes emotional conditions like severe depression) of Selected Person
39	Speech impairment of Selected Person prevented participation, no translator available
40	Language problem – Selected Person's language not understood by survey team, no translator available
	Individual modules 2-7 begun but not completed
51	Refusal by Selected Person to continue modules 2-7
52	Unusual circumstance - Selected Person
	General Booklet (module 9) not begun
61	Unable to read or write in language of General Booklet – Selected Person refused to begin
62	Refusal by Selected Person to begin General Booklet
63	Unusual circumstance - Selected Person
64	Blindness or visual impairment of Selected Person prevented completion of General Booklet
65	Physical disability of Selected Person prevented completion of General Booklet
	General booklet (module 9) looked through or attempted but not completed
71	Selected Person looked at Booklet but unable to read & write the language of Booklet - did not begin
72	Refusal by Selected Person to continue the General Booklet
73	Interruption – unacceptably long
74	Unusual circumstance - Selected Person
75	Attempted General Booklet but no answers marked
76	Marked some items but did not attempt all items in General Booklet
77	Marked some items & attempted all items in General Booklet
78	Marked all items in General Booklet
98	Sample selection error – an incorrect eligible person was selected by the interviewer
99	Sample selection error – an ineligible person was selected by the interviewer

Ineligible – A sampled household '**x**' is categorized as <u>ineligible</u> if there is no household member in the eligible 15-64 age range. The result codes for ineligible sampled households '**x**' are as follows:

II. Household-level Codes – Ineligible Household	
Code	**Code Description**
07	Dwelling could not be found / selected address has no household
08	Vacant dwelling
09	Dwelling not habitable/dwelling destroyed/dwelling converted to commercial use
31	No household member in the eligible range of 15-64

Non-participant - Eligibility Known: A non-participating household '**z**' results when an eligible household does not participate due to refusal, unusual circumstance, or the household was selected in error. The result codes for sampled households '**z**' with known eligibility are as follows:

III. Household-level Codes – Non-participants, Eligible Household	
Code	**Code Description**
21	Household Refusal to continue Household Module (i.e., begun but not complete)
22	Household unusual circumstance (e.g., death in family, illness, fire in dwelling, etc.) prevented continuance
97	Sample selection error – a wrong household was selected in the field

Non-participant - Eligibility Unknown: A non-participating household '**u**' results when a household does not participate due to non-contact with a responsible household member, inconclusive information about the eligibility status of the sampled household, or sample case was not activated (this usually pertains to the reserve sample cases). The result codes for non-participating households '**u**' with unknown eligibility are as follows:

IV. Household-level Codes – Non-participants, Unknown Eligibility Household	
Code	**Code Description**
01	Household Refusal
02	Household unusual circumstance (e.g., death in family, illness, fire in dwelling, etc.)
03	No knowledgeable household member found after 3 visits (only a child, non-competent adult, etc.)
04	Temporarily absent/unavailable for field period (information from other persons)
05	No competent household member to interview (because of severe illness, mental disability, etc.)
06	Language problem – no household member spoke a language understandable by survey team, no translator available
95	Sample case not activated
96	General non-response; Reason unspecified by survey firm

The total number of sampled households in each PSU may be represented by the following equation.

$$5) \qquad n_j = n_{jk} + n_{jx} + n_{jz} + n_{ju}$$

where,
n_j - the number of households sampled in the jth PSU, (i.e., n_j =30 households in each China-Yunnan sampled PSU),
n_{jk} - the number of households 'k' that participate in the jth PSU,
n_{jx} - the number of ineligible households 'x' in the jth PSU,
n_{jz} - the number of non-participating households 'z' with known eligibility in the jth PSU,
n_{ju} - the number of non-participating households 'u' with unknown eligibility in the jth PSU.

Table 10 provides the STEP counts from the China-Yunnan data file by the outcome categories described above.

Table 10: Sample Count of Households by Final Response Category		
Final Response Category	Frequency	Percent
1-Participant HH (n_{jk})	2020	49.9
2-Ineligible HH (n_{jx})	8	.2
3-Non-participant, Eligible HH (n_{jz})	5	.1
4-Non-participant, Unknown Eligibility HH (n_{ju})	2017	49.8
Total (n_j)	4050	100.0

5.2.2 Adjustment of Household Weights for Unknown Eligibility

The STEP assumes that the households 'u' with unknown eligibility in the jth PSU are comprised of some proportion, say ε, that are eligible households and the complementary proportion, $1 - \varepsilon$, that are ineligible households.

Amongst the n_{ju} non-participating households 'u' with unknown eligibility the proportion of households with known eligibility is estimated to be

$$6) \qquad \varepsilon = \frac{Total\ \#\ of\ known\ eligible\ cases}{Total\ \#\ of\ known\ eligible\ \&\ known\ ineligible\ cases} = \frac{n_{jk} + n_{jz}}{n_{jk} + n_{jx} + n_{jz}}$$

➤ **Note that the eligible proportion is the same for all eligible participating households.**

Example 3: Eligible Proportion

The counts in Table 10 provide input for calculation of the proportion of households with known eligibility.

$$\varepsilon = \frac{Total\ \#\ of\ known\ eligible\ cases}{Total\ \#\ of\ known\ eligible\ \&\ known\ ineligible\ cases}$$

$$= \frac{n_{jk} + n_{jz}}{n_{jk} + n_{jx} + n_{jz}}$$

$$= \frac{2020 + 5}{2020 + 8 + 5} = \frac{2025}{2033} \approx 0.996065$$

5.2.3 Adjustment of Household Weights for Non-participation

A household-level response adjustment is necessary to compensate for households that did not participate in the survey. The household-level response adjustment for the jth PSU, R_j, is calculated as follows:

$$7) \qquad R_j = \frac{Total\ eligible\ households\ in\ PSU\ 'j'}{Total\ participating\ households\ in\ PSU\ 'j'} = \frac{n_{jk} + n_{jz} + (\varepsilon * n_{ju})}{n_{jk}}$$

➤ **The response adjustment ratio, R_j, incorporates the adjustment for unknown eligibility, ε.**
➤ **The households that are categorized as ineligible, n_{jx} , are excluded from the calculation of the household-level response adjustment.**

<div style="border:1px solid #000;">

Example 4: Second Stage Response Adjustment for sampled Households

For example, the following results were obtained for PSU #126

n_j	n_{jk}	n_{jx}	n_{jz}	n_{ju}
30	15	0	0	15

Therefore,

$$R_{126} = \frac{Total\ eligible\ households\ in\ PSU\ j}{Total\ particpating\ households\ in\ PSU\ j} = \frac{n_{jk} + n_{jz} + (\varepsilon * n_{ju})}{n_{jk}}$$
$$= \frac{15 + 0 + (.99065 * 15)}{15}$$
$$\approx 1.996065$$

</div>

The response adjustment ratio, R_j , is similarly calculated for all 135 sampled PSUs.

5.2.4 Final Second-stage Household Weight

The final household weight is the weight used for producing household-level estimates, such as the proportion of households with a specific attribute. This weight is appended to the STEP data file.

The final household weight is calculated as the product of the PSU weight, the basic household weight, and the response adjustment for the basic household weight. Therefore, for household 'k' in PSU 'j', the final household weight, W_H (i.e., weighted data file variable '**W_FinHHwt**'),is calculated as follows:

8) $W_H = W_{1j} * W_{2jk} * R_j$

➢ **The final second stage household weight is appended to the STEP data file.**

<div style="border:1px solid #000;">

Example 5: Second Stage – Final Household Weight

For PSU #126,

$$W_H = W_{1j} * W_{2jk} * R_j$$
$$= 43.92 \times 6.67 \times 1.966065$$
$$\approx 584.4478$$

</div>

The household weight is similarly calculated for all sampled households.

5.2.5 Weighted Household Counts

The original sample frame 'sampling frame for STEP _CHINA' showed that there is a total of 1,067,256 households in the target population.

For the 135 sampled PSUs in the China-Yunnan data file, with the exception of PSU #126, the number of households in each PSU listing was the same as the number of households shown in the sample frame. The data file of STEP results indicates that there are 200 listed households in PSU #126, while the Sample Frame indicated 180 households in this PSU. This increase of 20 households results in an increase in the weighted estimate of the number of households in the target population.

Table 11 provides,
  i.   the weighted count of households based on the sample frame count of households for the 135 sampled PSUs;
 ii.   the weighted count of households based on the China data file indication of the number of listed households.

Table 11: China-Yunnan – Weighted Number of Households – Sample Frame vs. STEP Data File		
Source of PSU Household Count	# of Sampled PSUs	Weighted Sum of Households
Based on # of Hhlds in Sample Frame 'sampling frame for STEP _CHINA'	135	1,067,256
Based on # of Listed Hhlds in STEP Data File 'ch_rst_code' [**all participants**, i.e., participants, non-participants, ineligibles]	135	1,068,134
Based on # of Listed Hhlds in STEP Data File 'ch_rst_code' [**eligible households only**]	135	1,063,933

Note that the increase in the weighted number of households based on the listed household counts can be attributed to the increased count in PSU #126, the only PSU with a different listed count of households from the sample frame count of households.

Recall that the approximate final household weight for PSU #126 is 584.4478. Since there are an additional 20 listed households in PSU #126 the estimated additional households based on the household listings is

$$Additional\ Listed\ HHlds\ =\ \text{PSU weight for PSU #126}\ \times \text{# of additional listed household}$$
$$=\ 43.92 * 20$$
$$\approx\ 878$$

Therefore, the revised weighted number of household count based on the listed households for the 135 sampled PSUs is

$$Wtd\ Listed\ HHlds\ =\ \text{Wtd Frame count of Hhlds}\ +\ Additional\ Wtd\ Listed\ HHlds$$
$$=\ 1,067,256\ +\ 878$$
$$\approx\ 1,068,134$$

This weighted total number of households, 1068134, includes all households categorized as participants, non-participants, and ineligible.

As shown in the third row of Table 11, when ineligible households are excluded, the weighted number of eligible households is 1,063,933.

5.3 Third-stage Weight (Person Weight)

The basic third stage weight for the ith person, W_{3jki} , is calculated as follows:

$$9) \qquad W_{3jki} = \frac{a_{jk}}{b_{jk}} = \frac{a_{jk}}{1} = a_{jk}$$

where,

   a_{jk} - the number of eligible persons in the kth household in the jth PSU,

   b_{jk}- the number of persons sampled in the kth household in the jth PSU, (i.e., n_{jk} =1 person in each China-Yunnan sampled household).

   N.B. $\qquad A_{jk} = \sum_{k=1}^{n_{jk}} a_{jk} = total\ \#\ of\ selected\ persons\ in\ the\ j^{th}\ PSU$

   where,

   n_{jk} - the number of households 'k' that participate in the jth PSU

Example 6: Third Stage – Basic Person Weight within sampled Household

For PSU #126, selected household # 02 has 3 eligible persons. Therefore,

$$W_{3jki} = \frac{a_{jk}}{b_{jk}} = \frac{3}{1} = 3$$

5.3.1 Third-stage Person-level Response Adjustment

For eligible participating households, there are three possible person-level outcomes, a selected eligible person participates in the survey, a selected eligible person does not participate, a selected person is ineligible.

Participant SP - A participating selected person '**i**' is an eligible sampled person that provides sufficient data to be considered a participant in the survey. The sampled persons '**i**' with result codes 51-78 satisfy the requirement for classification of the person as a survey participant. The result codes for participating persons '**i**' are as follows:

I.	Person-level Codes – Participant in Eligible Selected Household
Code	**Code Description**
	Individual modules 2-7 begun but not completed
51	Refusal by Selected Person to continue modules 2-7
52	Unusual circumstance - Selected Person
	General Booklet (module 9) not begun
61	Unable to read or write in language of General Booklet – Selected Person refused to begin
62	Refusal by Selected Person to begin General Booklet
63	Unusual circumstance - Selected Person
64	Blindness or visual impairment of Selected Person prevented completion of General Booklet
65	Physical disability of Selected Person prevented completion of General Booklet
	General booklet (module 9) looked through or attempted but not completed
71	Selected Person looked at Booklet but unable to read & write the language of Booklet - did not begin
72	Refusal by Selected Person to continue the General Booklet
73	Interruption – unacceptably long
74	Unusual circumstance - Selected Person
75	Attempted General Booklet but no answers marked
76	Marked some items but did not attempt all items in General Booklet
77	Marked some items & attempted all items in General Booklet
78	Marked all items in General Booklet

Non-participant SP - A non-participating person '**q**' results when an eligible person does not participate due to refusal, unusual circumstance, non-contact, language problem, physical disability (hearing, vision, other) or the person was selected in error. The result codes for non-participating persons '**q**' are as follows:

II.	Person-level Codes – Non-participant in Eligible Selected Household
Code	**Code Description**
	Individual modules 2-7 not begun
32	Refusal by Selected Person to begin modules 2-7
33	Unusual circumstance - Selected Person (e.g., death in family, illness, fire in dwelling, etc.)
34	Refusal - Another household member refused to let selected individual do individual modules
35	Unable to contact Selected Person after three visits during field period
36	Temporarily absence/unavailability of Selected Person for field period (info from household member)
37	Hearing Problem - Selected Person is deaf or so hard-of-hearing, no translator available
38	Learning or mental disability (includes emotional conditions like severe depression) of Selected Person
39	Speech impairment of Selected Person prevented participation, no translator available
40	Language problem – Selected Person's language not understood by survey team, no translator available
98	Sample selection error – an incorrect eligible person was selected by the interviewer

Ineligible SP - A sampled person '**f**' is categorized as <u>ineligible</u> if the person is not a member of the target population, for example, the person's age is outside the eligible 15-64 age range. Ineligible sampled persons '**f**' are as assigned the result code 99.

III. Person-level Codes – Non-participant in Eligible Selected Household	
Code	**Code Description**
99	Sample selection error – an ineligible person was selected by the interviewer

The third stage person weight is only relevant for sampled households in the final response category FinRespCat=1, i.e., participating households. Table 12 provides the pertinent frequency counts of the Final Result Codes.

Table 12: Final Result Code for Participating Households (i.e., Households with FinRespCat_HH=1)

Final Result Code & Description	Frequency	Percent
52-Interview termination-SP unusual circumstance	1	.0
62-SP refusal to begin GB	19	.9
64-SP blindness or visual impairment-unable to do GB	1	.0
72-SP refused to continue GB after starting	4	.2
73-GB interruption	2	.1
74-GB terminated due to unusual circumstance	4	.2
76-SP marked some GB items-did not attempt all items	30	1.5
77-SP marked some GB items & attempted all GB items	69	3.4
78-SP marked all GB items	1887	93.4
98-Sample selection error – incorrect eligible person selected by the interviewer	2	.1
99-Sample selection error – ineligible person selected by the interviewer	1	.0
Total	**2020**	**100.0**

In each participating household in the STEP Survey, one eligible person was selected to be interviewed. Therefore, for each participating household the basic person weight, W_{3jki} , would also be the final person weight within a sampled PSU <u>if all persons in the PSU participated</u>. However, when there is one or more non-participating selected persons within a sampled PSU, the basic person-weights for the PSU require a response adjustment so that the final person weight for each participating person in the PSU is also representative of the non-participating person(s) in the sampled households in the PSU. In other words, the total number of participating persons in a sampled PSU must represent the total number of eligible persons in the sampled households in the PSU.

Therefore, within each sampled PSU, a person-level response adjustment to the basic person weight is necessary to compensate for persons that did not participate in the survey.

The total number of eligible persons in each PSU may be represented by the following equation:

$$10)\ A_{jk} = A_{jki} + A_{jkq}$$

where,

$A_{jk} = \sum_{k=1}^{n_{jk}} a_{jk}$ -   the total number of eligible persons for all n_{jk} participating households in the jth PSU,

$A_{jki} = \sum_{k=1}^{n_{jk}} a_{jki}$ -   the total number of eligible persons 'i' that participate for all n_{jk} participating households in the jth PSU,

$A_{jkq} = \sum_{k=1}^{n_{jk}} a_{jkq}$ -   the total number of non-participating eligible persons 'q' for all n_{jk} participating households in the jth PSU,

Therefore, the person-level response adjustment for the jth PSU, R_{jk}, is calculated as follows:

$$11)\ \boldsymbol{R_{jk}} = \frac{\textit{Total eligible persons in all participating sampled households in PSU 'j'}}{\textit{Total particpating persons in all participating sampled households in PSU 'j'}} = \frac{A_{jk}}{A_{jki}}$$

Note that ineligible persons are excluded from the above ratio adjustment for person-level non-response.

Example 7: Third Stage – Basic Person Weight Response Adjustment

i) For PSU #126, there are 44 eligible persons across all participating households. Also, all selected persons participated in the survey and thus the selected persons represent the 44 eligible persons in the participating households. Therefore,

$$A_{jk} = \sum_{k=1}^{n_{jk}} a_{jk} = 44\ ; \qquad A_{jki} = \sum_{k=1}^{n_{jk}} a_{jki} = 44$$

$$\boldsymbol{R_{jk}} = \frac{A_{jk}}{A_{jki}} = \frac{\boldsymbol{44}}{\boldsymbol{44}} = \boldsymbol{1.00}$$

ii) For PSU # 212, the total number of eligible persons across all participating households is 40. For PSU # 212, there is one non-participating selected person from a sampled household with 3 eligible persons. Hence, the total number of eligible persons across the participating households where the selected person participates is 37 (i.e., 40 minus 3). Therefore,

$$A_{jk} = \sum_{k=1}^{n_{jk}} a_{jk} = 40\ ; \qquad A_{jki} = \sum_{k=1}^{n_{jk}} a_{jki} = 37$$

$$\boldsymbol{R_{jk}} = \frac{A_{jk}}{A_{jki}} = \frac{\boldsymbol{40}}{\boldsymbol{37}} = \boldsymbol{1.08}$$

Table 13 provides the sample counts of the Final Response Category for selected persons.

Table 13: Sample Counts by Selected Person Final Response Category for Participating Households

For Participating Households		
Final Response Category-SP	**Frequency**	**Percent**
1-Participant SP	2017	49.8
2-Non-participant, Eligible	2	.0
3-Non-participant, Ineligible	1	.0
9-Not applicable - non-participating HH	2030	50.1
Total	**4050**	**100.0**

Based on the above table, there are 2017 participating selected persons for which a final non-zero person weight was calculated. The sampled persons in categories 2, 3, and 9 will have a zero value final person weight.

5.3.2 Final Third-stage Person Weight

The final person weight is the weight used for producing person-level estimates, such as the proportion of persons with a specific attribute. This weight is appended to the STEP data file.

The final person weight is calculated as the product of the final household weight (which incorporates the PSU weight and the response adjustment for the basic household weight), the basic third stage person weight, and the response adjustment for the basic person weight. Therefore, for the selected person 'i' in household 'k' in PSU 'j', the final person weight, W_{SP} (i.e., weighted data file variable '**W_FinSPwt**'), is calculated as follows:

$$12)\ W_{SP} = W_H * W_{3jki} * R_{jk}$$

➢ **The final third stage person weight is appended to the STEP data file.**

Example 8: Third Stage – Final Person Weight

i) For PSU #126, selected household # 02 with 3 eligible persons, the selected person was a participant. All selected persons participated in PSU #126, therefore, $R_{jk} = 1$

$$W_{SP} = 584.4478 * 3 * 1 \approx 1753.34$$

ii) For PSU #212, in participant household # 07 with 5 eligible persons, the selected person was a participant. For this household,

$$W_H \approx 526.003 ; \qquad W_{3jki} \approx 5 ; \qquad R_{jk} \approx 1.08$$

$$W_{SP} = 526.003 * 5 * 1.08 \approx 2630.02$$

The final person weight is similarly calculated for all sampled persons. Note that the final person weight will be zero for all non-participating selected persons.

5.3.3 Weighted Person Counts

The weighted count of the non-institutionalized persons 15 to 64 years of age (inclusive) living in private dwellings in urban areas of **Kunming** at the time of data collection is provided in Table 14. Based on the 2017 participating selected persons, the sum of the final person weights from the STEP survey is 2,612,811.

Table 14: Sample Frequency & Weighted Persons Totals by Final Selected Person Response Category		
For Participating Households		
Final Response Category-Selected Person	**# of Sampled Cases**	**Weighted Total # of Eligible Persons in Target Population**
1-Participant SP	2,017	2,612,811
2-Non-participant, Eligible	2	0
3- Non-participant, Ineligible	1	0
9-Not applicable - non-participating HH	2,030	0
Total	**4,050**	**2,612,811**

For participating selected persons, Table 15 provides a summary of the central tendency of the number of eligible persons in the participating households. Based on the average of 2.45 eligible persons per participating household and an estimated 1,063,933 households, there are approximately 2,600,000 persons aged 15 to 64 in the target population.

Table 15: Summary of Average # of Eligible Persons, Average Person Weights & Average Household Weights			
Statistic	**NUMBER A- # of eligible persons**	**Final Person Weight**	**Final Household Weight**
	2017	2017	2017
Mean	2.45	1289.72	526.69
Std. Error of Mean	0.018	9.66	0.16
Median	2.00	1052.01	526.00
Mode	2	1052.01	526.00
Minimum	1	508.50	508.50
Maximum	6	3156.02	584.45
Sum	**4938**	**2,612,811**	**1,063,933**