

October 11, 2013

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

**Bolivia**

**October 11, 2013**



## Table of Contents

1	Survey Design Overview.....	1
2	Data Collection Outcomes .....	3
3	Weighting Overview.....	5
4	Weighting Process Files.....	6
4.1	Revisions to Input Data File .....	6
4.2	Final Output File from Weighting Process .....	10
5	Weight Calculations based on World Bank (WB) Weight Requirement .....	14
5.1	PSU Weight .....	14
5.2	Household Weight .....	17
5.2.1	Household Weight - Response Adjustment .....	17
5.2.2	Household Weight - Adjustment for Unknown Eligibility .....	21
5.2.3	Household Weight - Adjustment for Non-participation .....	21
5.2.4	Household Weight - Final.....	22
5.2.5	Weighted Household Counts .....	23
5.3	Person Weight.....	25
5.3.1	Person Weight - Person-level Response Adjustment .....	26
5.3.2	Person Weight - Activated Eligible Sample Cases .....	30
5.3.3	Person Weight - Final.....	31
5.3.4	Weighted Person Counts .....	32

October 11, 2013

## 1 Survey Design Overview

The Bolivia survey firm implemented a Full literacy assessment design. The Full assessment required each selected person to attempt to complete a General Booklet comprising Reading Components and a set of Core Literacy Items. Subsequently, those respondents that correctly answered at least 3 Core Literacy items were assigned one of four assessment Exercise Booklets to complete. The full assessment sampling objective was to have a minimum of about 2400 completed Exercise Booklets, with a distribution of about 600 complete cases for each of the four Exercise Booklets.

The STEP target population comprises all non-institutionalized persons 15 to 64 years of age (inclusive) living in private dwellings in urban areas of the main capital cities of the country at the time of data collection. The cities that are covered are La Paz, El Alto, Cochabamba and Santa Cruz de la Sierra.

### Exclusions

Urban areas in small cities are excluded; these urban exclusions comprise approximately 20% of the urban target population.

The sample frame for the selection of first stage sample units was the Stata file '[sample frame at psu level](#)' that is based on the Bolivia survey firm's data and was provided by the World Bank core team. The frame is a list of first stage sampling units in the target population. The source of the sample frame of the first stage units is the 2001 National Census of Population and Housing carried out by the National Institute of Statistics.

The Bolivia sample design is a stratified 3 stage sample design. The stratification variable is city-wealth category. There are 20 strata created by grouping the primary sample units (PSUs) into the 4 cities, i.e., 1- La Paz, 2-El Alto, 3-Cochabamba, 4-Santa Cruz de la Sierra, and 5 wealth categories, i.e., 1-Poorest, 2-Moderately Poor, 3-Middle Wealth, 4-Moderately Rich, 5-Rich.

Table 1 provides the sample frame counts by city and wealth category.

Wealth Category	La Paz		El Alto		Cochabamba		Santa Cruz de la Sierra		Total	
	PSUs	HHlds	PSUs	HHlds	PSUs	HHlds	PSUs	HHlds	PSUs	HHlds
1	45	2504	295	19,013	43	2,860	56	4,180	439	28,557
2	1,198	85403	2,033	132,233	519	36,554	635	46,984	4,385	301,174
3	1,307	97,987	188	12,881	725	54,565	2,167	166,728	4,387	332,161
4	129	8,525	5	384	246	18,552	203	14,195	583	41,656
5	145	9,671	2	123	143	10,860	220	15,623	510	36,277
<b>Total</b>	<b>2,824</b>	<b>204,090</b>	<b>2,523</b>	<b>164,634</b>	<b>1,676</b>	<b>123,391</b>	<b>3,281</b>	<b>247,710</b>	<b>10,304</b>	<b>739,825</b>

### First Stage Sample

The primary sample unit (PSU) is a Census Sector. A sample of 218 PSUs was selected from the 10,304 PSUs on the sample frame. This sample of PSUs was comprised of 160 'initial' PSUs and 58 'reserve' PSUs.

**Note:** Of the 218 sampled PSUs, there were 169 activated PSUs consisting of 155 Initial Sampled PSUs and 14 Reserve sampled PSUs. Among the 160 'initial' PSUs, 5 PSUs were replaced due to security concerns; also, 14 reserve PSUs were activated to supplement the sample for initial PSUs where the target sample of 15 interviews was not achieved due to high levels of non-response; thus, only 169 PSUs were actually activated during data collection.

The PSUs were grouped according to city-wealth strata, and within each city-wealth stratum PSUs were selected with probability proportional to size (PPS), where the measure of size was the number of households in a PSU.

### Second Stage Sample

The second stage sample unit (SSU) is a household. The sampling objective was to obtain interviews at 15 households within each of the initial PSU sample, resulting in a final initial sample of 2400 interviews. At the second stage of sample selection, 45 households were selected in each PSU using a systematic random method. The 45 households were randomly divided into 15 'Initial' households, and 30 'Reserve' households that were ranked according to the random sample selection order.

**Note:** Due to higher than expected levels of non-response in some PSUs, additional households were sampled; thus, the final actual sample in some PSUs exceeded 45 households.

### 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.

Table 2 provides the sample counts by city-wealth strata.

**Table 2: Bolivia Sample Counts – PSUs & Households by City-Wealth Stratum**

Wealth Indicator- PSU	Number of Activated Sampled PSUs					Number of Sampled Households				
	(includes Initial & Reserve PSUs)					(includes Initial & Reserve Households)				
	City				Total	City				Total
	1-LPZ	2-ALTO	3-CBB	4-SCZ		1-LPZ	2-ALTO	3-CBB	4-SCZ	
1-Poor	5	10	4	6	25	225	450	180	316	1171
2-Modestly Poor	14	20	6	12	52	680	900	270	544	2394
3-Middle	16	7	8	18	49	821	322	361	849	2353
4-Modestly Rich	10	2	5	10	27	483	90	232	469	1274
5-Rich	4	1	4	7	16	197	45	180	387	809
<b>Total</b>	<b>49</b>	<b>40</b>	<b>27</b>	<b>53</b>	<b>169</b>	<b>2406</b>	<b>1807</b>	<b>1223</b>	<b>2565</b>	<b>8001</b>

## 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 final result code assigned to the sampled household and the final result code assigned to the selected person. The result codes used in the Bolivia STEP are listed in Table 3; a ‘final\_result\_code’ was assigned to each sampled case.

<b>Table 3: Full Literacy Assessment Final Result Codes for Sampled Households and Selected Persons</b>	
<b>I. Household-level Codes</b>	
<b>Code</b>	<b>Code Description</b>
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
10	Household moved away
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 field procedure error – wrong household selected in its place
<b>II. Person-level Codes</b>	
<b>Code</b>	<b>Code Description</b>
<b>Individual modules 2-7 not begun or unacceptable interview</b>	
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
<b>Individual modules 2-7 begun but not completed</b>	
51	Refusal by Selected Person to continue modules 2-7
52	Unusual circumstance - Selected Person
<b>General Booklet (module 9) not begun</b>	
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

<b>Table 3: Full Literacy Assessment Final Result Codes for Sampled Households and Selected Persons</b>	
65	Physical disability of Selected Person prevented completion of General Booklet
	<b>General Booklet (module 9) looked through or attempted but not completed and not passed</b>
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	Attempted part or all of General Booklet; less than 3 correct Core answers
	<b>Exercise Booklet (module 9)</b>
81	No attempt at any of the exercise booklet items / refusal to start
82	Interruption too long during exercise booklet administration
83	Unusual circumstance, Exercise Booklet administration terminated
84	Attempt at some Exercise Booklet items but no answers marked
85	Attempt at all Exercise Booklet items but no answers marked
86	Marked some Exercise Booklet items, not all items attempted
87	Marked some Exercise Booklet items, all items attempted
88	Marked all Exercise Booklet items

### 3 Weighting Overview

The Bolivia stratified three-stage cluster design 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, the Bolivia STEP requires a sampling weight for each person that participated in the survey.

In general, the objectives of the Bolivia 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 Bolivia STEP survey requires 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, and an Exercise Booklet, and all household and individual questionnaire modules.

- a) Calculation of a PSU weight for 169 activated sample 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 Bolivia data file with the survey design weights (i.e., for each sampled PSU, household, person) appended to each data record.

## 4 Weighting Process Files

The key file for the weighting process is the STATA file '[STEP - Bolivia - Data for Weights - Revised Feb 1 2013 - Revised Feb 1 2013](#)' provided by the WB core team. An SPSS version of this file was created for the weighting process.

### 4.1 Revisions to Input Data File

The following revisions were made to the Bolivia data file '[STEP - Bolivia - Data for Weights - Revised Feb 1 2013 - Revised Feb 1 2013](#)':

- 1) Revised the value labels for the variables '*m11\_q07*', the Result of Interview Code, and '*final\_result\_code*', the final status code of a sampled case, to correspond to the Codes and code descriptions in Table 3 of this document.
- 2) Revised the Final Result Code, i.e., variable '*final\_result\_code*', as follows:
  - ◆ RECODE *final\_result\_code* (95=1) for PSUs 105, 108, 110, 143, 178, 203, 210, 248, 261.
  - ◆ IF (cluster = 103 & hhn=33) *final\_result\_code*=95.
  - ◆ IF (cluster = 120 & (hhn=54 | hhn=55)) *final\_result\_code*=95.
  - ◆ IF (cluster = 122 & (hhn=35 | hhn=28)) *final\_result\_code*=95.
  - ◆ IF (cluster = 153 & (hhn=1 | hhn=10 | hhn=15)) *final\_result\_code*=95.
  - ◆ IF (cluster = 188 & hhn=25) *final\_result\_code*=95.
  - ◆ IF (cluster = 191 & (hhn=14 | hhn=1 | hhn=16)) *final\_result\_code*=95.
  - ◆ IF (cluster = 201 & hhn=31) *final\_result\_code*=95.
  - ◆ IF (cluster = 215 & (hhn=23 | hhn=9)) *final\_result\_code*=95.
  - ◆ IF (cluster = 284 & hhn=14) *final\_result\_code*=95.
  - ◆ IF (cluster = 292 & (hhn=3 | hhn=28)) *final\_result\_code*=95.
  - ◆ IF (cluster = 304 & hhn=49) *final\_result\_code*=95.
  - ◆ IF (cluster = 179 & (hhn=45 | hhn=43 | hhn=3)) *final\_result\_code*=97.
  - ◆ IF (cluster = 204 & (hhn=39 | hhn=42 | hhn=44 | hhn=45 )) *final\_result\_code*=97.
  - ◆ IF (cluster = 205 & (hhn=23 | hhn=39 | hhn=40 | hhn=41 | hhn=42 | hhn=43 | hhn=44 | hhn=45)) *final\_result\_code*=97.
  - ◆ IF (cluster = 147 & hhn=34) *final\_status*=98.
  - ◆ IF (cluster = 164 & hhn=25) *final\_status*=98.
  - ◆ IF (cluster = 166 & hhn=15) *final\_status*=98.
  - ◆ IF (cluster = 163 & hhn=40) *final\_status*=98.
  - ◆ IF (cluster = 233 & hhn=14) *final\_status*=99.
  - ◆ IF (cluster = 158 & hhn=4) *final\_status*=99.
- 3) Missing Values
  - i) Replaced missing values with the correct value for the variables HhldsPSU, PopnPSU, city, wealth, hhsmpyp, PSUsmp\_Type, HH\_Replaced\_Ind, PSUsmpId\_Ind, country.
  - ii) Variable '*hhsmpord*' – Replaced '*hhsmpord*' missing value with a sequential value beginning with '*hhsmpord*'=31.
- 4) Created the variable '*FinStat1*', a revised final status code that distinguishes non-activated cases that should have been activated from non-activated cases that should not have been activated, i.e., these latter cases were activated in an order that did not respect the prescribed activation order that was based on the sampling order of the reserve sample cases.
  - a) With the exception of code 95, the values for '*FinStat1*' are the same as the code values for '*final\_status*'.

- b) In the case of 'final\_status'=95, the values for 'FinStat1' are as follows,
- i) Code 95 - a reserve household was not activated and should not have been activated.
  - ii) Code 97 - a household should have been activated, but was not activated during the survey period due to an out-of-order household that was selected in its place.

**Note:** The 'final status' codes 95 & 97 are treated the same way in the weighting process – they are both treated as non-participant households (eligibility unknown). So, the calculation of the sample weights is not affected by the revised 'FinStat1' status codes. However, the revision of the 'final status' codes to 'FinStat1' is relevant to the calculation of the response rate and to any subsequent non-response bias analysis and sample selection bias analysis that may be undertaken.

5) Created the variable 'Stratum\_CityWlth', city-wealth stratum as follows:

- i) IF (city=1 & wealth= 1) Stratum\_CityWlth=11.
- ii) IF (city=1 & wealth= 2) Stratum\_CityWlth=12.
- iii) IF (city=1 & wealth= 3) Stratum\_CityWlth=13.
- iv) IF (city=1 & wealth= 4) Stratum\_CityWlth=14.
- v) IF (city=1 & wealth= 5) Stratum\_CityWlth=15.
- vi) IF (city=2 & wealth= 1) Stratum\_CityWlth=21.
- vii) IF (city=2 & wealth= 2) Stratum\_CityWlth=22.
- viii) IF (city=2 & wealth= 3) Stratum\_CityWlth=23.
- ix) IF (city=2 & wealth= 4) Stratum\_CityWlth=24.
- x) IF (city=2 & wealth= 5) Stratum\_CityWlth=25.
- xi) IF (city=3 & wealth= 1) Stratum\_CityWlth=31.
- xii) IF (city=3 & wealth= 2) Stratum\_CityWlth=32.
- xiii) IF (city=3 & wealth= 3) Stratum\_CityWlth=33.
- xiv) IF (city=3 & wealth= 4) Stratum\_CityWlth=34.
- xv) IF (city=3 & wealth= 5) Stratum\_CityWlth=35.
- xvi) IF (city=4 & wealth= 1) Stratum\_CityWlth=41.
- xvii) IF (city=4 & wealth= 2) Stratum\_CityWlth=42.
- xviii) IF (city=4 & wealth= 3) Stratum\_CityWlth=43.
- xix) IF (city=4 & wealth= 4) Stratum\_CityWlth=44.
- xx) IF (city=4 & wealth= 5) Stratum\_CityWlth=45

**Value Labels added for 'Stratum\_CityWlth'**

11-LPZ-Poor	21-ALTO-Poor	31-CBB-Poor	41-SCZ-Poor
12-LPZ-Modestly Poor	22-ALTO-Modestly Poor	32-CBB-Modestly Poor	42-SCZ-Modestly Poor
13-LPZ-Middle	23-ALTO-Middle	33-CBB-Middle	43-SCZ-Middle
14-LPZ-Modestly Rich	24-ALTO-Modestly Rich	34-CBB-Modestly Rich	44-SCZ-Modestly Rich
15-LPZ-Rich	25-ALTO-Rich	35-CBB-Rich	45-SCZ-Rich

6) **SPSS Output File: 'W01-BOL Final Revised Data File(4-Feb-2013)'**

- a) This file includes all 8001 sampled cases; each case includes a final result code.
- i) Table 4 provides the breakdown of the 8001 cases by Final Status Code & type of household sample.
    - ◆ For weighting purposes, a finalized case is a sample household for which a final result code has been assigned. A final result code was assigned to the 8001 sampled cases.
  - ii) This file is the primary file for the process of creating the survey weights.

- ♦ Table 5 provides the list of key variables for weighting that are included in 'W01-BOL Final Revised Data File(4-Feb-2013)'

**Table 4: Bolivia Step Data - Number of Sample Cases by Final Status Code by**

<b>Final Status Code (FinStat1 - Revised for out-of-order cases)</b>	<b>Initial HH Sample</b>	<b>Reserve HH Sample</b>	<b>Total</b>
01-Refusal-Household	762	1566	2328
02-Unusual circumstance-Household	5	9	14
03-No knowledgeable household member found after 3 visits	121	208	329
04-Temporarily absent/unavailable for field period	44	65	109
05-No competent household member (e.g., severe illness, mental disability, etc.)	1	3	4
06-Language problem	6	9	15
07-Dwelling could not be found / no household at selected address	15	24	39
08-Vacant dwelling	176	308	484
09-Dwelling not habitable/destroyed/converted to commercial	40	69	109
21-Household Refusal to continue Household Module (begun,not complete)	21	37	58
22-Household module terminated - unusual circumstance	2	3	5
31-No eligible household member aged 15-64	50	60	110
32-Refusal by SP to begin modules 2-7	126	123	249
33-Interview termination-SP unusual circumstance	5	4	9
34-Refusal-Other Household Member refused to let SP participate	17	18	35
35-Unable to contact SP after 3 attempts	43	39	82
36-SP absent for entire survey period	6	8	14
37-SP deaf or hard of hearing	1	0	1
38-SP learning or mental disability	1	3	4
40-SP language problem	0	1	1
51-Refusal by SP to continue modules 2-7	10	13	23
52-Interview termination-SP unusual circumstance	1	3	4
61-SP unable to read language of GB	15	7	22
62-SP refusal to begin GB	13	16	29
63-SP unable to do GB due to unusual circumstance	0	1	1
64-SP blindness or visual impairment-unable to do GB	1	2	3
71-SP looked through GB-unable to read/write language of GB	2	6	8
72-SP refused to continue GB after starting	30	23	53
75-Attempted General Booklet, no answers marked	3	4	7
76-Attempted part or all of General Booklet; <3 correct Core answers	139	180	319
81-No attempt at any exercise booklet items / refusal to start	9	18	27
83-Unusual circumstance, Exercise Booklet administration terminated	2	1	3
84-Attempt at some Exercise Booklet items, no answers marked	2	4	6
85-Attempt at all Exercise Booklet items, no answers marked	4	2	6
86-Marked some Exercise Booklet items, not all items attempted	30	26	56
87-Marked some Exercise Booklet items, all items attempted	373	474	847
88-Marked all Exercise Booklet items	446	575	1021
95-Household not activated	0	1518	1518
96-General non-response; unspecified reason	0	1	1
97-Household not interviewed - field procedure error	6	30	36
98-Incorrect eligible household member selected	7	3	10
99-Ineligible household member selected	0	2	2

**Table 4: Bolivia Step Data - Number of Sample Cases by Final Status Code by**

Final Status Code (FinStat1 - Revised for out-of-order cases)	Initial HH Sample	Reserve HH Sample	Total
<b>Total</b>	<b>2535</b>	<b>5466</b>	<b>8001</b>

**Table 5: Key Variables in Primary Input File for Weighting Process**

Variable Name	Source	Description
<b>country</b>	Sample File	Country abbreviation code
<b>cluster</b>	Bolivia Data File	Household Id
<b>psusmpid</b>	Sample File	PSU ID – sampled PSUs
<b>hhn</b>	Bolivia Data File	Household Number
<b>city</b>	Bolivia Data File	City
<b>wealth</b>	Bolivia Data File	Wealth category for PSU
<b>HhldsPSU</b>	Sample File (from Sample Frame)	Sample Frame # of households in PSU
<b>HHsmpSeq</b>	Sample File	Household Sample Sequence #s 01 to 30
<b>HHsmpType</b>	Sample File	HH Sample Indicator
<b>HHsmpOrder</b>	Sample File	Sample activation order
<b>m1c_a</b>	Bolivia Data File	NUMBER A-# of eligible persons
<b>m1c_b</b>	Bolivia Data File	NUMBER B- listing # of selected person
<b>m11_q07</b>	Bolivia Data File	Result code of interview
<b>final_result_code</b>	Bolivia Data File	Final Result Code
<b>HHcount_Listd</b>	Bolivia Data File	total # of listed households in PSU
<b>Stratum_CityWlth</b>	Created during file preparations	City-wealth Stratum

## 4.2 Final Output File from Weighting Process

The output file from the weighting process is,

- 1) [BOL-W01-STEP\(HH\) Weighted Data-Basic Weights \(4-Feb-2013\)](#)
  - a) This file contains the original data variables from the Bolivia data file '[STEP - Bolivia - Data for Weights - Revised Feb 1 2013 - Revised Feb 1 2013](#)' and the applicable PSU weights, household weights, person weights, and other variables that were created during the weighting process.
  - b) The file contains 8001 data records, of which there are 2842 participating households with a final household weight. The breakdown of the count of participating households by urban-rural indicator is shown below in Table 6.

<b>City- Wealth Stratum</b>	<b>Participating Eligible Sample Households</b>	<b>Weighted Number of Households</b>
11-LPZ-Poor	71	2,717
12-LPZ-Modestly Poor	263	74,855
13-LPZ-Middle	299	100,429
14-LPZ-Modestly Rich	128	11,434
15-LPZ-Rich	67	12,595
21-ALTO-Poor	171	37,421
22-ALTO-Modestly Poor	327	143,293
23-ALTO-Middle	129	13,184
24-ALTO-Modestly Rich	36	419
25-ALTO-Rich	16	42
31-CBB-Poor	48	3,972
32-CBB-Modestly Poor	99	27,460
33-CBB-Middle	134	44,601
34-CBB-Modestly Rich	85	13,623
35-CBB-Rich	66	10,513
41-SCZ-Poor	104	7,277
42-SCZ-Modestly Poor	215	47,147
43-SCZ-Middle	310	160,243
44-SCZ-Modestly Rich	172	14,750
45-SCZ-Rich	102	16,004
<b>Total</b>	<b>2842</b>	<b>741,978</b>

- c) In addition to the Bolivia data file variables, the weighted data file includes 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, created or revised during the weighting process, which are included in the final weighted data file.
  - ii) Table 8 provides the list of other available 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 '[BOL-W01-STEP\(HH\) Weighted Data-Basic Weights \(4-Feb-2013\)](#)'.

**Table 7: Variables Created or Revised During Weighting Process - Added to Bolivia Data File**

<b>Variable Name</b>	<b>Description</b>	<b>Values</b>
<b>HhldsPSU</b>	Sample Frame # of households in PSU (renamed from 'househol')	n/a
<b>HHcount_Listd</b>	total # of listed households in PSU (renamed from 'hhcount')	n/a
<b>Stratum_CityWlth</b>	City Wealth Category	'11-LPZ-Poor' '12-LPZ-Modestly Poor' '13-LPZ-Middle' '14-LPZ-Modestly Rich' '15-LPZ-Rich' '21-ALTO-Poor' '22-ALTO-Modestly Poor' '23-ALTO-Middle' '24-ALTO-Modestly Rich' '25-ALTO-Rich' '31-CBB-Poor' '32-CBB-Modestly Poor' '33-CBB-Middle' '34-CBB-Modestly Rich' '35-CBB-Rich' '41-SCZ-Poor' '42-SCZ-Modestly Poor' '43-SCZ-Middle' '44-SCZ-Modestly Rich' '45-SCZ-Rich'
<b>final_status</b>	Final Result Code	Refer to Table 2 for code descriptions
<b>FinStat1</b>	Final Status-Revised for out-of-order cases	Refer to Table 2 for code descriptions
<b>FinRespCat_HH</b>	Final Response Category-Households	1-Participant HH 2-Ineligible HH 3-Non-participant Eligible HH 4-Non-participant Unknown Eligibility HH 9-HH Missing Final Status Code
<b>FinRespCat_SP</b>	Final Response Category-Selected Person	1-Participant SP 2-Non-participant, Eligible 3-Non-participant, Ineligible 7-Non-participant SP-non-participant activated HH 8-Not applicable - non-activated HH 9-Not applicable - Ineligible HH
<b>W1_PSUwt</b>	Basic PSU Weight	n/a
<b>W_FinHHwt</b>	Final Household Weight Use for estimation of Household characteristics	n/a
<b>W_FinSPwt</b>	Final Person Weight Use for estimation of Person characteristics	n/a

**Table 8: Variables Created During Weighting Process – Not Added to Bolivia Final Weighted Data File**

<b>Variable Name</b>	<b>Description</b>
<b>hhsmpsiz</b>	Number of Sampled HHs in PSU
<b>PSU_FIN#</b>	Count of 'FinRespCat_HH' in PSU
<b>HH_k</b>	# of participant HHs in PSU
<b>HH_x</b>	# of Ineligible HHs in PSU
<b>HH_z</b>	# of Non-participant Eligible HHs in PSU
<b>HH_u</b>	# of Non-participant Unknown Eligible HHs in PSU
<b>HH_k_StratTot</b>	# of participant HHs in Stratum
<b>HH_x_StratTot</b>	# of Ineligible HHs in Stratum
<b>HH_z_StratTot</b>	# of Non-participant Eligible HHs in Stratum
<b>HH_u_StratTot</b>	# of Non-participant Unknown Eligible HHs in Stratum
<b>W2_HH</b>	Household Weight within PSU
<b>W2_HHwt_basic</b>	Basic Household Weight
<b>Elig_Prop</b>	Proportion of Eligible HHs - known eligibility
<b>R_adj_HH</b>	Response adjustment ratio for HHs
<b>EligPersCnt</b>	Eligible Person Count in Sampled HH
<b>W3_SP</b>	Selected Person n weight within household
<b>R_adj_SP</b>	Response adjustment for selected person
<b>EP_PSU_sum</b>	Total Eligible Persons-PSU Participant Households
<b>EP_PSU_sum_PP</b>	Total Eligible Persons-PSU Participant SP Households

**Table 9: Bolivia Final Weighted Data File – Frequency of Final Result Codes For participating households (i.e., FINRESPCAT\_HH=1)**

Final Result Code	Total	
	n	% of Total n
32-Refusal-Refusal by SP to begin modules 2-7	249	8.8
33-Interview termination-SP unusual circumstance	9	.3
34-Refusal-Other Household Member refused to let SP participate	35	1.2
35-Unable to contact SP after 3 attempts	82	2.9
36-SP deaf or hard of hearing	14	.5
37-SP speech impairment	1	.0
38-SP learning or mental disability	4	.1
40-SP language problem	1	.0
51-Refusal by SP to continue modules 2-7	23	.8
52-Interview termination-SP unusual circumstance	4	.1
61-SP unable to read language of GB	22	.8
62-SP refusal to begin GB	29	1.0
63-SP unable to do GB due to unusual circumstance	1	.0
64-SP blindness or visual impairment-unable to do GB	3	.1
71-SP looked through GB-unable to read & write language of GB	8	.3
72-SP refused to continue GB after starting	53	1.9
75-SP attempted GB-did not mark any answers	7	.2
76-SP marked some GB items-did not attempt all items	319	11.2
81-No attempt at any of the exercise booklet items / refusal to start	27	1.0
83-Unusual circumstance, Exercise Booklet administration terminated	3	.1
84-Attempt at some Exercise Booklet items but no answers marked	6	.2
85-Attempt at all Exercise Booklet items but no answers marked	6	.2
86-Marked some Exercise Booklet items, not all items attempted	56	2.0
87-Marked some Exercise Booklet items, all items attempted	847	29.8
88-Marked all Exercise Booklet items	1021	35.9
98-Sample selection error – incorrect eligible person selected	10	.4
99-Ineligible household member selected	2	.1
<b>Total</b>	<b>2842</b>	<b>100.0%</b>

## 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 '[W01-BOL Final Revised Data File\(4-Feb-2013\)](#)', which was created from the WB provided file '[STEP - Bolivia - Data for Weights - Revised Feb 1 2013 - Revised Feb 1 2013](#)'.

### 5.1 PSU Weight

The basic first stage weight, i.e., the PSU weight, is calculated separately for each selected stratum. The stratum PSU weight is based on an overall sample size of 169 urban PSUs (155 Initial PSUs and 14 Reserve PSUs). Although 218 PSUs were initially selected, 174 PSUs were attempted during data collection; 5 initial PSUs were replaced due to safety concerns for the field staff, resulting in only 169 PSUs actually activated during data collection. Therefore, the Bolivia STEP PSU survey weight is based on the probability of selection for each of the 169 activated PSUs.

For the Bolivia PPS sample, the probability of selection of the  $j^{\text{th}}$  PSU in city-wealth stratum 'g', say  $p_{gj}$ , may be represented as,

$$1) \quad p_{gj} = \frac{y_g * m_{gj}}{M_g}$$

where,

$y_g$  - the number of sampled PSUs in stratum g,

$m_{gj}$  - the measure of size for the  $j^{\text{th}}$  PSU in stratum 'g', in the case of the Bolivia STEP Survey the measure of size for the first stage selection of PSUs was the sample frame number of households in each PSU in the target population,

$M_g$  - the total number of target population households in stratum 'g', based on the sample frame information, is represented by

$$2) \quad M_g = \sum_1^{y_g} m_{gj}$$

Table 10 provides the Bolivia stratum counts for  $y_g$  and  $M_g$ .

Table 10: Bolivia – Activated Sample PSUs & Sample Frame Households by Stratum				
City	Wealth Category	Stratum	Sampled PSUs	Frame Households
		$g$	$y_g$	$M_g$
1-La Paz	1-Poor	1	5	2,504
	2-Moderate Poor	2	14	85,403
	3-Middle Wealth	3	16	97,987
	4-Moderate Rich	4	10	8,525
	5-Rich	5	4	9,671
	Total La Paz			49
2-El Alto	1-Poor	6	10	19,013
	2-Moderate Poor	7	20	132,233
	3-Middle Wealth	8	7	12,881
	4-Moderate Rich	9	2	384
	5-Rich	10	1	123
	Total El Alto			40
3-Cochabamba	1-Poor	11	4	2,860
	2-Moderate Poor	12	6	36,554
	3-Middle Wealth	13	8	54,565
	4-Moderate Rich	14	5	18,552
	5-Rich	15	4	10,860
	Total Cochabamba			27
4-Santa Cruz de la Sierra	1-Poor	16	6	4,180
	2-Moderate Poor	17	12	46,984
	3-Middle Wealth	18	18	166,728
	4-Moderate Rich	19	10	14,195
	5-Rich	20	7	15,623
	Total Santa Cruz de la Sierra			53
Total	1-Poor		25	28,557
	2-Moderate Poor		52	301,174
	3-Middle Wealth		49	332,161
	4-Moderate Rich		27	41,656
	5-Rich		16	36,277
	Total			169

The basic first stage weight,  $W_{1gj}$ , for the  $j^{\text{th}}$  PSU in stratum 'g' is calculated as the inverse of its probability of selection:

$$3) \quad W_{1gj} = \frac{1}{p_{gj}} = \frac{M_g}{y_g * m_{gj}}$$

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

**Example 1: First Stage Weight**

The basic first stage weight for Bolivia PSU # 101 in stratum 3 – ‘La Paz Middle Wealth’, which contains 71 sample frame households, is calculated as follows:

$$W_{13j} = \frac{1}{p_{3j}} = \frac{M_g}{y_1 * m_{1j}} = \frac{97,987}{16 * 71} \approx 86.26$$

The basic first stage weight is similarly calculated for each of the 169 PSUs.

## 5.2 Household Weight

The basic second stage weight for the  $k^{\text{th}}$  household in PSU 'j',  $W_{2gjk}$ , is calculated as follows:

$$4) \quad W_{2gjk} = \frac{m'_{gj}}{n_{gj}}$$

where,

$m'_{gj}$  - the number of **listed** households in the  $j^{\text{th}}$  PSU in stratum g,

$n_{gj}$  - the number of households sampled in the  $j^{\text{th}}$  PSU in stratum g.

### Example 2: Second Stage Basic Weight

The basic second stage weight for each of the 45 sampled households in the Bolivia PSU # 101, which contains 54 **listed** households, is calculated as follows:

$$W_{2gjk} = \frac{m'_{gj}}{n_{gj}} = \frac{54}{45} \approx 1.2$$

The basic second stage weight is similarly calculated for each of the 169 PSUs.

### 5.2.1 Household Weight - 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 Household** - A participating household 'k' is an eligible 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, 51-88, 94 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:

#### I. Household-level Codes – Participant, Eligible Household

Code	Code Description
<b>Individual modules 2-7 not begun</b>	
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
<b>Individual modules 2-7 begun but not completed</b>	
51	Refusal by Selected Person to continue modules 2-7
52	Unusual circumstance - Selected Person
<b>General Booklet (module 9) not begun</b>	
61	Unable to read or write in language of General Booklet – Selected Person refused to begin

### I. Household-level Codes – Participant, Eligible Household

Code	Code Description
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
<b>General booklet (module 9) looked through or attempted but not completed</b>	
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
81	No attempt at any of the exercise booklet items / refusal to start
82	Interruption too long during exercise booklet administration
83	Unusual circumstance, Exercise Booklet administration terminated
84	Attempt at some Exercise Booklet items but no answers marked
85	Attempt at all Exercise Booklet items but no answers marked
86	Marked some Exercise Booklet items, not all items attempted
87	Marked some Exercise Booklet items, all items attempted
88	Marked all Exercise Booklet items
94	Case dropped due to data quality issue
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 Household** – A sampled household ‘x’ is categorized as ineligible 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
10	Household moved away
31	No household member in the eligible range of 15-64

**Non-participant Household - 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

**Non-participant Household - 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:

<b>IV. Household-level Codes – Non-participants, Unknown Eligibility Household</b>	
<b>Code</b>	<b>Code Description</b>
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
97	Household not interviewed due to field procedure error – wrong household selected in its place

The total number of sampled households in stratum ‘g’ may be represented by the following equation.

$$5) \quad n_g = n_{gk} + n_{gx} + n_{gz} + n_{gu}$$

where,

- $n_g$  - the number of households sampled in stratum ‘g’,
- $n_{gk}$  - the number of households ‘k’ that participate in stratum ‘g’,
- $n_{gx}$  - the number of ineligible households ‘x’ in stratum ‘g’,
- $n_{gz}$  - the number of non-participating households ‘z’ with known eligibility in stratum ‘g’,
- $n_{gu}$  - the number of non-participating households ‘u’ with unknown eligibility in stratum ‘g’.

Table 11 provides the STEP counts from the Bolivia data file for the outcome categories described above.

<b>Table 11: Bolivia Sample Count of Households by Final Response Category by Stratum</b>					
	<b>Final Response Category-Households</b>				
<b>Stratum-City Wealth Category</b>	<b>1-Participant</b>	<b>2-Ineligible</b>	<b>3-Non-participant Eligible</b>	<b>4-Non-participant Unknown Eligibility</b>	<b>Total</b>
<b>(g)</b>	<b>(n<sub>gk</sub>)</b>	<b>(n<sub>gx</sub>)</b>	<b>(n<sub>gz</sub>)</b>	<b>(n<sub>gu</sub>)</b>	
1-LPZ-Poor	71	9	1	144	225
2-LPZ-Modestly Poor	263	63	5	349	680
3-LPZ-Middle	299	100	5	417	821
4-LPZ-Modestly Rich	128	54	1	300	483
5-LPZ-Rich	67	19		111	197
6-ALTO-Poor	171	53	2	224	450
7-ALTO-Modestly Poor	327	80	9	484	900
8-ALTO-Middle	129	30	5	158	322
9-ALTO-Modestly Rich	36	10	1	43	90
10-ALTO-Rich	16	7		22	45
11-CBB-Poor	48	4		128	180
12-CBB-Modestly Poor	99	22	1	148	270
13-CBB-Middle	134	33	5	189	361
14-CBB-Modestly Rich	85	33	3	111	232
15-CBB-Rich	66	22		92	180
16-SCZ-Poor	104	17	2	193	316
17-SCZ-Modestly Poor	215	52	7	270	544
18-SCZ-Middle	310	57	5	477	849
19-SCZ-Modestly Rich	172	39	8	250	469
20-SCZ-Rich	102	38	3	244	387
<b>Total</b>	<b>2842</b>	<b>742</b>	<b>63</b>	<b>4354</b>	<b>8001</b>

## 5.2.2 Household Weight - Adjustment for Unknown Eligibility

The STEP assumes that the households 'u' with unknown eligibility in the j<sup>th</sup> PSU in stratum 'g' are comprised of some proportion, say  $\varepsilon_g$ , that are eligible households and the complementary proportion,  $1 - \varepsilon_g$ , that are ineligible households. Note that the calculation of  $\varepsilon_g$  is carried out independently for the PSUs in each stratum.

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

$$6) \quad \varepsilon_g = \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} = \frac{n_{gk} + n_{gz}}{n_{gk} + n_{gx} + n_{gz}}$$

- **Note that the eligible proportion is the same for all eligible participating households in stratum 'g'.**

### **Example 3: Eligible Proportion**

The counts in Table 11 provide input for calculation of the proportion of households with known eligibility.

For the PSUs in stratum 3, 'La Paz-Middle Wealth', the calculation of the eligibility proportion is as follows,

$$\begin{aligned} \varepsilon_1 &= \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} \\ &= \frac{n_{1k} + n_{1z}}{n_{1k} + n_{1x} + n_{1z}} \\ &= \frac{299 + 5}{299 + 100 + 5} = \frac{304}{404} \approx 0.75275 \end{aligned}$$

## 5.2.3 Household Weight - Adjustment for Non-participation

A household-level response adjustment is necessary to compensate for households that did not participate in the survey.

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

$$7) \quad n_j = n_{jk} + n_{jx} + n_{jz} + n_{ju}$$

where,

- $n_j$  - the number of households sampled in PSU 'j',
- $n_{jk}$  - the number of households 'k' that participate in PSU 'j',
- $n_{jx}$  - the number of ineligible households 'x' in PSU 'j',
- $n_{jz}$  - the number of non-participating households 'z' with known eligibility in PSU 'j',
- $n_{ju}$  - the number of non-participating households 'u' with unknown eligibility in PSU 'j'.

The household-level response adjustment,  $R_{gj}$ , for the  $j^{\text{th}}$  PSU in stratum 'g' is calculated as follows:

$$8) \quad R_{gj} = \frac{\text{Total eligible households in PSU } 'j'}{\text{Total participating households in PSU } 'j'} = \frac{n_{jk} + n_{jz} + (\varepsilon_g * n_{ju})}{n_{jk}}$$

- The response adjustment ratio,  $R_j$ , incorporates the adjustment for unknown eligibility,  $\varepsilon_g$ .
- The households that are categorized as ineligible,  $n_{jk}$ , are excluded from the calculation of the household-level response adjustment.

**Example 4: Second Stage Response Adjustment for sampled Households**

For example, the following results were obtained for PSU #101

$n_j$	$n_{jk}$	$n_{jx}$	$n_{jz}$	$n_{ju}$
45	15	3	4	23

Therefore,

$$\begin{aligned} R_{1,101} &= \frac{\text{Total eligible households in PSU } j}{\text{Total participating households in PSU } j} = \frac{n_{jk} + n_{jz} + (\varepsilon_1 * n_{ju})}{n_{jk}} \\ &= \frac{17 + 2 + (0.752475 * 23)}{17} \\ &\approx 2.1357 \end{aligned}$$

The response adjustment ratio,  $R_j$ , is similarly calculated for each activated sample PSU.

### 5.2.4 Household Weight - Final

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,  $W_H$ , is calculated as follows:

$$9) \quad W_H = W_{1gj} * W_{2gjk} * R_{gj}$$

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

**Example 5: Second Stage – Final Household Weight**

For PSU #101,

$$\begin{aligned} W_H &= W_{1j} * W_{2jk} * R_j \\ &= 86.26 \times 1.2 \times 2.1357 \\ &\approx 221.06 \end{aligned}$$

The final household weight is similarly calculated for all sampled households.

## 5.2.5 Weighted Household Counts

The original sample frame ‘[sample frame at psu level](#)’ showed that there is a total of 739,825 households in the target population. The numbers of households on the sample frame includes households that no longer existed at the time of the STEP survey, e.g. vacant dwellings, demolished dwellings, no eligible person for STEP in household, etc..

The total un-weighted number of households in the sampled PSUs is provided in Table 12. The counts in Table 12 are un-weighted counts that have not been adjusted for ineligible households found in the STEP sample. This table illustrates the difference in the PSU listed household counts versus the PSU household counts from the sample frame; overall, there are about 29% more households from the PSU field listing than the number of PSU households indicated on the sample frame.

**Table 12: Bolivia Number of Households in Sampled PSUs by Stratum & Source of Count**

Stratum-City Wealth Category	PSUs	Households	
	Activated	Sample Frame	Field Listing
11-LPZ-Poor	5	325	356
12-LPZ-Modestly Poor	14	1043	1152
13-LPZ-Middle	16	1070	1416
14-LPZ-Modestly Rich	10	529	951
15-LPZ-Rich	4	248	390
21-ALTO-Poor	10	594	974
22-ALTO-Modestly Poor	20	1294	1711
23-ALTO-Middle	7	530	651
24-ALTO-Modestly Rich	2	154	214
25-ALTO-Rich	1	84	41
31-CBB-Poor	4	260	399
32-CBB-Modestly Poor	6	422	373
33-CBB-Middle	8	598	585
34-CBB-Modestly Rich	5	341	342
35-CBB-Rich	4	288	368
41-SCZ-Poor	6	498	1053
42-SCZ-Modestly Poor	12	921	1160
43-SCZ-Middle	18	1427	1560
44-SCZ-Modestly Rich	10	714	894
45-SCZ-Rich	7	538	727
<b>Total</b>	<b>169</b>	<b>11,878</b>	<b>15,317</b>

Table 13 provides,

- i. the total un-weighted number of households based on the sample frame count of households;
- ii. the weighted count of households based on the listed households from the Bolivia listing operation.

**Table 13: Bolivia –Weighted Number of Households – Sample Frame vs. STEP Data File**

Stratum-City Wealth Category	Number of Households		
	Un-Weighted	Weighted	Weighted
		(All sampled cases)	(Eligible cases only)
	Sample Frame	Field Listing	Field Listing
(1)	(2)	(3)	
11-LPZ-Poor	2,504	3,129	2,717
12-LPZ-Modestly Poor	85,403	92,229	74,855
13-LPZ-Middle	97,987	133,492	100,429
14-LPZ-Modestly Rich	8,525	16,256	11,434
15-LPZ-Rich	9,671	16,186	12,595
21-ALTO-Poor	19,013	48,737	37,421
22-ALTO-Modestly Poor	132,233	176,573	143,293
23-ALTO-Middle	12,881	16,130	13,184
24-ALTO-Modestly Rich	384	531	419
25-ALTO-Rich	123	60	42
31-CBB-Poor	2,860	4,318	3,972
32-CBB-Modestly Poor	36,554	33,302	27,460
33-CBB-Middle	54,565	55,293	44,601
34-CBB-Modestly Rich	18,552	18,542	13,623
35-CBB-Rich	10,860	13,793	10,513
41-SCZ-Poor	4,180	8,511	7,277
42-SCZ-Modestly Poor	46,984	58,376	47,147
43-SCZ-Middle	166,728	189,457	160,243
44-SCZ-Modestly Rich	14,195	17,915	14,750
45-SCZ-Rich	15,623	21,904	16,004
<b>Total</b>	<b>739,825</b>	<b>924,733</b>	<b>741,978</b>

Notes for Table 13

1. For column 1, the ‘Number of households’ is an un-weighted total; it is the actual total number of households for all PSUs in the sample frame.
2. For column 2, the increase in the weighted number of households based on the listed household counts can be attributed to the increased count of households from the listing operation for the sampled PSUs. These weighted estimates of the number of households are not adjusted for sampled ineligible households.
3. For column 3, the decrease from the column 2 weighted number of households is attributable to the adjustment of the household weights for ineligible households found in the STEP sample.
  - a. *In this case, the ‘Weighted Number of Households’ is the estimate of the number of households in the STEP target population.*

### 5.3 Person Weight

The basic third stage weight for the  $i^{\text{th}}$  person,  $W_{3jki}$ , is calculated as follows:

$$10) \quad W_{3jki} = \frac{a_{jk}}{b_{jk}} = \frac{a_{jk}}{1} = a_{jk}$$

where,

$a_{jk}$  - the number of eligible persons in the  $k^{\text{th}}$  household in the  $j^{\text{th}}$  PSU,

$b_{jk}$  - the number of persons sampled in the  $k^{\text{th}}$  household in the  $j^{\text{th}}$  PSU, (i.e.,  $n_{jk} = 1$  person in each Bolivia sampled household).

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

$n_{jk}$  - the number of households 'k' that participate in the  $j^{\text{th}}$  PSU

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

For PSU #101, selected household # 16 has 6 eligible persons. Therefore,

$$W_{3jki} = \frac{a_{jk}}{b_{jk}} = \frac{6}{1} = 6$$

### 5.3.1 Person Weight - 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, or 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 51-88 satisfy the requirement for classification of the person as a survey participant. The result codes for participating persons ‘i’ are as follows:

<b>I. Person-level Codes – Participant in Eligible Selected Household</b>	
<b>Code</b>	<b>Code Description</b>
<b>Individual modules 2-7 begun but not completed</b>	
51	Refusal by Selected Person to continue modules 2-7
52	Unusual circumstance - Selected Person
<b>General Booklet (module 9) not begun</b>	
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
<b>General booklet (module 9) looked through or attempted but not completed</b>	
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
81	No attempt at any of the exercise booklet items / refusal to start
82	Interruption too long during exercise booklet administration
83	Unusual circumstance, Exercise Booklet administration terminated
84	Attempt at some Exercise Booklet items but no answers marked
85	Attempt at all Exercise Booklet items but no answers marked
86	Marked some Exercise Booklet items, not all items attempted
87	Marked some Exercise Booklet items, all items attempted
88	Marked all Exercise Booklet items

**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:

<b>II. Person-level Codes – Non-participant in Eligible Selected Household</b>	
<b>Code</b>	<b>Code Description</b>
<b>Individual modules 2-7 <u>not</u> begun</b>	
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
94	Case dropped due to data quality issue
98	Sample selection error – an incorrect eligible person was selected by the interviewer

**Ineligible SP** - A sampled person ‘f’ is categorized as ineligible 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 assigned the result code 99.

<b>III. Person-level Codes – Non-participant in Eligible Selected Household</b>	
<b>Code</b>	<b>Code Description</b>
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\_HH=1, i.e., participating households. Table 14 provides the pertinent frequency counts of the Final Result Codes.

**Table 14: Final Result Code for Participating Households (i.e., Households with FinRespCat\_HH=1)**

Final Result Code & Description	Frequency	Percent
32-Refusal by Selected Person to begin modules 2-7	249	8.8
33-Interview termination-SP unusual circumstance	9	.3
34-Refusal-Other Household Member refused to let SP participate	35	1.2
35-Unable to contact SP after 3 attempts	82	2.9
36-SP absent for entire survey period	14	.5
37-SP deaf or hard of hearing	1	.0
38-SP learning or mental disability	4	.1
40- SP language problem	1	.0
51-Refusal by SP to continue modules 2-7	23	.8
52-Interview termination-SP unusual circumstance	4	.1
61- Unable to read or write in language of General Booklet – SP refused to begin	22	.8
62-SP refusal to begin GB	29	1.0
63- Unusual circumstance - Selected Person	1	.0
64-SP blindness or visual impairment-unable to do GB	3	.1
71-SP looked through GB-unable to read & write language of GB	8	.3
72-SP refused to continue GB after starting	53	1.9
75-Attempted General Booklet but no answers marked	7	.2
76-Attempted part or all of General Booklet; less than 3 correct Core answers	319	11.2
81-No attempt at any of the exercise booklet items / refusal to start	27	1.0
83-Unusual circumstance, Exercise Booklet administration terminated	3	.1
84-Attempt at some Exercise Booklet items but no answers marked	6	.2
85-Attempt at all Exercise Booklet items but no answers marked	6	.2
86-Marked some Exercise Booklet items, not all items attempted	56	2.0
87-Marked some Exercise Booklet items, all items attempted	847	29.8
88-Marked all Exercise Booklet items	1021	35.9
98-Sample selection error – incorrect eligible person selected	10	.4
99-Ineligible household member selected	2	.1
<b>Total</b>	<b>2842</b>	<b>100.0</b>

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 if all persons in the PSU participated. 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:

$$11) \quad A_{jk} = A_{jki} + A_{jkq}$$

where,

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

Therefore, the person-level response adjustment for the  $j^{\text{th}}$  PSU,  $R_{jk}$ , is calculated as follows:

$$12) \quad R_{jk} = \frac{\text{Total eligible persons in all participating sampled households in PSU } 'j'}{\text{Total participating persons in all participating sampled households in PSU } 'j'} = \frac{A_{jk}}{A_{jki}}$$

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**

For PSU #101, there are 51 eligible persons across all participating households. For PSU # 101, the total number of eligible persons across the participating households where a selected person participated is 47. Therefore,

$$A_{jk} = \sum_{k=1}^{n_{jk}} a_{jk} = 51 ; \quad A_{jki} = \sum_{k=1}^{n_{jk}} a_{jki} = 47$$

$$R_{jk} = \frac{A_{jk}}{A_{jki}} = \frac{51}{47} = 1.085$$

### 5.3.2 Person Weight - Activated Eligible Sample Cases

Table 15 shows the number of sampled cases that were not activated (i.e., final\_status code≠95) and the number of ineligible cases (i.e., final\_status code =7, 8, 9, 10, 31) among the activated sample cases.

**Table 15: Bolivia - Number of Non-activated Sample Cases & Ineligible Cases by Final Result Code**

Final Result Code	Total
07-Dwelling could not be found / selected address has no household	39
08-Vacant dwelling	484
09-Dwelling not habitable/dwelling destroyed/dwelling converted to commercial use	109
31-No eligible household member aged 15-64	110
95-Reserve household not activated	1540
<b>Total</b>	<b>2282</b>

When the above cases are excluded there are 5719 eligible activated sample cases that were distributed by stratum. Table 16 provides the sample counts by stratum of the Final Response Category for selected persons in the 5719 activated eligible households.

**Table 16: Sample Counts (n) by Selected Person (SP) Final Response Category for Activated Eligible Households**

Stratum-City Wealth Category	1-Participant SP		2-Non-participant SP, Eligible		3-Non-participant SP, InEligible		7-Non-participant SP - non-participant activated HH		Total	
	n	%	n	%	n	%	n	%	n	%
11-LPZ-Poor	61	1.1%	10	0.2%	0	0.0%	108	1.9%	179	3.1%
12-LPZ-Modestly Poor	210	3.7%	53	0.9%	0	0.0%	260	4.5%	523	9.1%
13-LPZ-Middle	236	4.1%	63	1.1%	0	0.0%	347	6.1%	646	11.3%
14-LPZ-Modestly Rich	116	2.0%	12	0.2%	0	0.0%	274	4.8%	402	7.0%
15-LPZ-Rich	59	1.0%	8	0.1%	0	0.0%	91	1.6%	158	2.8%
21-ALTO-Poor	150	2.6%	20	0.3%	1	0.0%	144	2.5%	315	5.5%
22-ALTO-Modestly Poor	290	5.1%	37	0.6%	0	0.0%	274	4.8%	601	10.5%
23-ALTO-Middle	105	1.8%	24	0.4%	0	0.0%	109	1.9%	238	4.2%
24-ALTO-Modestly Rich	30	0.5%	6	0.1%	0	0.0%	24	0.4%	60	1.0%
25-ALTO-Rich	15	0.3%	1	0.0%	0	0.0%	16	0.3%	32	0.6%
31-CBB-Poor	46	0.8%	2	0.0%	0	0.0%	68	1.2%	116	2.0%
32-CBB-Modestly Poor	90	1.6%	9	0.2%	0	0.0%	59	1.0%	158	2.8%
33-CBB-Middle	121	2.1%	13	0.2%	0	0.0%	98	1.7%	232	4.1%
34-CBB-Modestly Rich	75	1.3%	10	0.2%	0	0.0%	64	1.1%	149	2.6%
35-CBB-Rich	60	1.0%	6	0.1%	0	0.0%	78	1.4%	144	2.5%
41-SCZ-Poor	90	1.6%	14	0.2%	0	0.0%	120	2.1%	224	3.9%
42-SCZ-Modestly Poor	180	3.1%	35	0.6%	0	0.0%	106	1.9%	321	5.6%
43-SCZ-Middle	267	4.7%	42	0.7%	1	0.0%	216	3.8%	526	9.2%
44-SCZ-Modestly Rich	141	2.5%	31	0.5%	0	0.0%	197	3.4%	369	6.5%
45-SCZ-Rich	93	1.6%	9	0.2%	0	0.0%	224	3.9%	326	5.7%
<b>Total</b>	<b>2435</b>	<b>42.6%</b>	<b>405</b>	<b>7.1%</b>	<b>2</b>	<b>0.0%</b>	<b>2877</b>	<b>50.3%</b>	<b>5719</b>	<b>100.0%</b>

Based on the above table, there are 2435 participating selected persons for which a final non-zero person weight was calculated. The sampled persons in categories 2 and 9 will have a zero value final person weight.

### 5.3.3 Person Weight - Final

**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,  $W_{SP}$  is calculated as follows:

$$13) \quad W_{SP} = W_H * W_{3jki} * R_{gjk}$$

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

#### **Example 8: Third Stage – Final Person Weight**

For PSU #101, selected household # 16 with 6 eligible persons, the selected person was a participant. For PSU #101, there are 51 eligible persons across all participating households. For PSU # 101, the total number of eligible persons across the participating households where a selected person participated is 47. Therefore,  $R_{jk} = 1.085$

$$W_{SP} = 221.06 * 6 * 1.085 \approx 1439.25$$

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

### 5.3.4 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 the main cities of **Bolivia** at the time of data collection is provided in Table 17. The 'Weighted # of Eligible Persons' in this table is the estimated number of persons aged 15 to 64 in the Bolivia target population based on the STEP sample.

**Table 17: Number of Participating Selected Persons & Weighted Total**

Stratum (City-Wealth)	Final Response Category - Selected Person	
	# of participating Eligible Persons	Weighted # of Eligible Persons
11-LPZ-Poor	61	10,511
12-LPZ-Modestly Poor	210	231,039
13-LPZ-Middle	236	289,585
14-LPZ-Modestly Rich	116	28,727
15-LPZ-Rich	59	31,313
21-ALTO-Poor	150	126,648
22-ALTO-Modestly Poor	290	452,974
23-ALTO-Middle	105	41,450
24-ALTO-Modestly Rich	30	1,352
25-ALTO-Rich	15	115
31-CBB-Poor	46	9,853
32-CBB-Modestly Poor	90	82,673
33-CBB-Middle	121	125,647
34-CBB-Modestly Rich	75	35,108
35-CBB-Rich	60	28,774
41-SCZ-Poor	90	20,444
42-SCZ-Modestly Poor	180	138,633
43-SCZ-Middle	267	445,967
44-SCZ-Modestly Rich	141	41,260
45-SCZ-Rich	93	48,758
<b>Total</b>	<b>2,435</b>	<b>2,190,831</b>

For participating selected persons, Table 18 provides a summary of the central tendency of the number of eligible persons in the participating households. Based on the average of 2.95 eligible persons per participating household and an estimated 741,978 households, there are approximately 2,190,000 persons aged 15 to 64 in the target population.

<b>Table 18: Summary of Average # of Eligible Persons, Average Person Weights &amp; Average Household Weights</b>			
<b>Statistic</b>	<b>NUMBER A- # of eligible persons</b>	<b>Final Person Weight</b>	<b>Final Household Weight</b>
	2836	2842	2842
<b>Mean</b>	2.95	770.88	261.08
<b>Std. Error of Mean</b>	.028	16.94	3.72
<b>Median</b>	3.00	483.93	221.06
<b>Mode</b>	2.00	0	576.88
<b>Minimum</b>	1.00	0	2.61
<b>Maximum</b>	12.00	7004.97	1714.26
<b>Sum</b>	8367	2,190,831	741,978