

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

Ghana

November 27, 2013

Table of Contents

1	Survey Design Overview.....	1
1.1	Target Population.....	1
1.2	Sample Frame	1
1.3	Sample Design.....	2
1.4	Sample Size	3
2	Data Collection Outcomes	4
3	Weighting Overview.....	6
4	Files - Weighting Process	7
4.1	Input Files.....	7
4.1.1	Revisions to Input Data File.....	7
4.1.1.1	Input File – Revised for Weighting	9
4.2	Final Output Files from Weighting Process.....	11
5	Weight Calculations based on World Bank (WB) Weight Requirement	15
5.1	First-stage Weight (Basic PSU Weight)	15
5.1.1	First-stage Response Adjustment	16
5.1.1.1	Eligible Proportion of PSUs	17
5.1.1.2	Response Adjustment Factor for Non-participation.....	17
5.1.2	Response-adjusted PSU Weight.....	17
5.2	Second-stage Weight (PSU Partition Weight).....	18
5.2.1	Final PSU Weight	18
5.3	Third-stage Weight (Household Weight)	19
5.3.1	Basic Household Weight	19
5.3.2	Third-stage Response Adjustment.....	20
5.3.2.1	Eligible Proportion of Households	23
5.3.2.2	Response Adjustment Factor for Non-participation.....	24
5.3.3	Final Third-stage Household Weight.....	25
5.4	Fourth-stage Weight (Person Weight)	26
5.4.1	Basic Fourth-stage Person Weight.....	26
5.4.2	Fourth-stage Person-level Response Adjustment.....	27
5.4.3	Final Fourth-stage Person Weight	31
6	Benchmark Weights – Adjustment of Weights to Known Population Totals	32
6.1	Benchmark-adjusted Household Weight	32

6.2	Benchmark-adjusted Person Weight	33
6.2.1	Person Weight Adjustment for Known Number of Households.....	33
6.2.2	Person Weight: Adjustment for Known Number of Persons, Gender, Age Group	34
7	Weighted Estimates from STEP.....	36
7.1	Weighted Household Count.....	36
7.2	Weighted Person Count	37
8	Response Rate.....	38

1 Survey Design Overview

The Ghana 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.

1.1 Target Population

The target population for the Ghana STEP survey comprises all non-institutionalized persons 15 to 64 years of age (inclusive) living in private dwellings in urban areas of the country at the time of data collection. This includes all residents except foreign diplomats and non-nationals working for international organizations.

Exclusions

Military barracks were excluded from the Ghana target population.

1.2 Sample Frame

The sample frame for the selection of first stage sample units was the SPSS file '[GHA Frame_Aug 7-2012](#)'. This file was created from the Excel file '[Frame_Aug](#)', which was provided by the Ghana survey firm. The source of the sample frame is the 2010 Ghana Population and Housing Census.

The sample frame included the population size for each urban PSU. Table 1 provides the sample frame count of the number of PSUs and the Population count by Region.

Table 1: Ghana Sample Frame - Number of PSUs & Population Size by Region

Region Code	Region Name	# of PSUs	Urban Population
01	Western	1,239	1,007,969
02	Central	1,350	1,037,878
03	Greater Accra	4,724	3,630,955
04	Volta	964	713,735
05	Eastern	1,708	1,143,918
06	Ashanti	3,618	2,897,290
07	Brong Ahafo	1,425	1,028,473
08	Northern	998	750,712
09	Upper East	324	219,646
10	Upper West	153	114,653
Total		16,503	12,545,229

1.3 Sample Design

The Ghana sample design is a four-stage sample design. There was no explicit stratification but the sample was implicitly stratified by Region. [Note: Implicit stratification was achieved by sorting the PSUs (i.e., *EACode*) by *RegnCode* and selecting a systematic sample of PSUs.]

First Stage Sample

The primary sample unit (PSU) was a Census Enumeration Area (EA).

Each PSU was uniquely defined by the sample frame variables *Regncode*, and *EACode*. The sample frame was sorted by *RegnCode* to implicitly stratify the sample frame PSUs by region. The sampling objective was to select 250 PSUs, comprised of 200 Initial PSUs and 50 Reserve PSUs. Although 250 PSUs were selected, only 201 PSUs were activated.

The PSUs were selected using a systematic probability proportional to size (PPS) sampling method, where the measure of size was the population size (i.e., *EAPopn*) in a PSU.

Second Stage Sample

The second stage sample unit is a PSU partition.

It was considered necessary to partition 'large' PSUs into smaller areas to facilitate the listing process.

The document '[Documentation on splitting of STEP EAs](#)' provided by the survey firm outlines the procedures for the partitioning of PSUs. The following rule was used to partition an EA:

Estimated Number of PSU Households *	Number of PSU Partitions
Less than 300	1
Between 300 and 500	2
More than 500	3

* [Estimated Number of PSU Households](#): According to the document '[Documentation on splitting of STEP EAs](#)', "the average number of households for each EA was estimated by dividing the population of the EA by the regional average household size that was recorded in the 2010 census from which the list of EAs was generated". The regional average household size is based on both urban and rural households; the use of the overall regional average would tend to underestimate the estimated number of households in a sampled PSU since rural localities tend to have higher average household sizes than urban localities.

* [Assumption](#): PSU partitions are equally sized, i.e., each PSU partition has the same Estimated Number of PSU Households.

After the partitioning of the PSUs, the survey firm randomly selected one partition. The selected partition was fully listed for subsequent enumeration in accordance with the field procedures.

Third Stage Sample

The third stage sample unit (SSU) is a household. The sampling objective was to obtain interviews at 15 households within each selected PSU. The households were selected in each PSU using a systematic random method.

Fourth Stage Sample

The fourth 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.

1.4 Sample Size

The Ghana firm's sampling objective was to obtain interviews from 3000 individuals in the urban areas of the country. In order to provide sufficient sample to allow for a worst case scenario of a 50% response rate the number of sampled cases was doubled in each selected PSU.

Although 50 extra PSUs were selected for use in case it was impossible to conduct any interviews in one or more initially selected PSUs only one reserve PSU was activated. Therefore, the Ghana firm conducted the STEP data collection in a total of 201 PSUs.

Table 2 provides the household sample size by Region for the activated PSUs.

Region	# of PSUs	Household Sample Size		
		Initial	Reserve	Total
01-Western	16	240	240	480
02-Central	17	255	255	510
03-Greater Accra	57	855	855	1710
04-Volta	12	180	180	360
05-Eastern	18	270	270	540
06-Ashanti	47	705	705	1410
07-Brong Ahafo	16	240	240	480
08-Northern	12	180	180	360
09-Upper East	4	60	60	120
10-Upper West	2	30	30	60
TOTAL	201	3015	3015	6030

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 3. A variable called 'final_status' with these code values was included in the Ghana final data file; a 'final_status' was assigned to each sampled case.

Table 3: <i>Full Literacy Assessment</i> Result Codes for Sampled Households and Selected Persons	
I. Household-level Codes	
Code	Code Description
01	Refusal-Household
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	07-Dwelling could not be found / selected address has no household
08	Vacant dwelling / selected address has no household
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 <u>not</u> 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

Table 3: <i>Full Literacy Assessment</i> Result Codes for Sampled Households and Selected Persons	
	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
76	Marked some items but did not attempt all items in General Booklet
	Exercise Booklet (module 9)
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

While the Ghana four-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 Ghana 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 201 activated 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 Ghana 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 file for the weighting process is the file '[STEP Ghana - Data for Weighting_S7.dta](#)' provided by the WB core team. The file '[STEP Ghana - Data for Weighting_S7.dta](#)' contains a data record for each of the 7,500 sampled households from the Ghana sample frame. An SPSS version of this file, i.e., '[STEP Ghana - Data for Weighting_S7.sav](#)', was created for the weighting process.

4.1.1 Revisions to Input Data File

The following general revisions were made to the Ghana SPSS data file '[STEP Ghana - Data for Weighting_S7](#)':

- 1) '[m11_q07](#)', the Result of Interview Code; revised as follows:
 - a) Added value labels; the value labels for '[m11_q07](#)' are shown in Table 4:

Table 4: VALUE LABELS for the variable 'm11_q07'	
00-No assigned code	51-Refusal by SP to continue modules 2-7'
01-Refusal-Household	52-Interview termination-SP unusual circumstance'
02-Unusual circumstance-Household	61-SP unable to read language of General Booklet (GB)'
03-No knowledgeable household member found after 3 visits	62-SP refusal to begin GB'
04-Temporarily absent/unavailable for field period	63-SP unable to do GB due to unusual circumstance'
05-No competent household member to interview (e.g., due to severe illness, mental disability, etc.)	64-SP blindness or visual impairment-unable to do GB'
06-Language problem	65-SP physical disability'
07-Dwelling could not be found / selected address has no household	71-SP looked through GB-unable to read & write language of GB'
08-Vacant dwelling / selected address has no household	72-SP refused to continue GB after starting'
09-Dwelling not habitable/dwelling destroyed/dwelling converted to commercial use	73-GB interruption too long'
10-household moved away	74-GB terminated due to unusual circumstance'
21-Household Refusal to continue Household Module (i.e., begun but not complete)'	75-Attempted GB but no answers marked'
22-Household module terminated due to unusual circumstance'	76-Attempted part or all of GB; less than 3 correct Core answers'
31-No eligible household member aged 15-64'	77-Marked some items & attempted all items in GB'
32-Refusal-Refusal by SP to begin modules 2-7'	78-Marked all items in GB'
33-Interview termination-SP unusual circumstance'	81-No attempt at any exercise booklet items / refusal to start'
34-Refusal-Other Household Member refused to let SP participate'	82-Interruption too long during exercise booklet administration'
35-Unable to contact SP after 3 attempts'	83-Unusual circumstance, Exercise Booklet administration terminated'
36-SP absent for entire survey period'	84-Attempt at some Exercise Booklet items, no answers marked'
37-SP deaf or hard of hearing'	85-Attempt at all Exercise Booklet items, no answers marked'
38-SP learning or mental disability'	86-Marked some Exercise Booklet items, not all items attempted'
39-SP speech impairment'	87-Marked some Exercise Booklet items, all items attempted'
40-SP language problem'	88-Marked all Exercise Booklet items'
	95-Household not activated'.

- b) Revised the values of 'm11_q07' as follows:
 - i) IF (MISSING(m11_q07)) m11_q07=0;
 - ii) IF (psusmpid = 234) m11_q07=31.
 - ♦ N.B. PSU 234 was an initially activated PSU that included only military barracks; these were excluded from the target population.

2) 'final_status_code', the Final status code; revised as follows:

- a) Renamed to '**final_status**'.
- b) '**final_status**' revised as follows:
 - i) IF (field_rst = 1 | field_rst = 3) Final_status=Final_status.
 - ii) IF (field_rst = 2) Final_status=95.
 - iii) IF (field_rst = 4) Final_status=97.
 - iv) IF (field_rst = 5) Final_status=98.
 - v) IF (field_rst = 6) Final_status=99.
 - vi) IF (psusmpid = 234) final_status=31.
- c) Refer to Table 3 in section 2 for the list of **final_status** codes.
 - i) The data file variable '*field_rst* - Fieldwork result code' was used to create the variable 'final_status'.
 - ♦ For '*field_rst*' the description of the codes is as follows:

Code	Code Description
1	Activated; for an 'initial sample' HH this means that it was contacted, for a 'reserve sample' HH it means that it was activated for an interview attempt.
2	Never activated; applicable to reserve sample HHs, the case was correctly never activated, i.e., the prescribed fieldwork procedures were followed.
3	Wrongly activated; applicable to reserve sample cases, the reserve case was activated, but according to the prescribed fieldwork procedures it should <u>not</u> have been activated (i.e., a reserve sample HH was incorrectly activated).
4	Wrongly never activated; applicable to reserve sample cases, the reserve case was never activated but according to the prescribed fieldwork procedures it should have been activated (i.e., a reserve HH should have been activated based on the provided activation order, but it wasn't).
5	Incorrect eligible HH member selected; during the random selection process, an incorrect household member was selected from the list of <u>eligible</u> members.
6	Ineligible HH member selected; during the random selection process, an <u>ineligible</u> household member was selected.

- 3) '*field_rst*' - Fieldwork result code; revised as follows:
 - i) IF (psusmpid = 234) field_rst=1.
- 4) '*PSU_HHs*' – Number of listed households in PSU
 - a) Added to the final data file; obtained from the Excel file '*STEP Ghana - Household Count per PSU*'.
- 5) '*PSU_parts*' - Number of parts into which PSU was split
 - a) Added to the final data file; obtained from the Excel file '*STEP Ghana - Household Count per PSU*'.

4.1.1.1 Input File – Revised for Weighting

The SPSS Revised Input File for weighting is ‘W00-GHA Final Revised Data File(20-Nov-2013)’. This file is the primary file for the process of creating the survey weights; it includes all 7500 sampled cases; each case includes a final status code. Table 5 provides the list of key variables for weighting purposes that are included in the file.

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#
PSUsmpID	Sample File	PSU Sample ID#
PSUsmp_Type	Sample File	PSU Sample Type (Initial, Reserve)
HHsmp_Order	Sample File	Sample activation order
HHsmpSeq	Sample File	Household Sample Sequence #s 01 to 30
HHsmp_Type	Sample File	HH Sample Indicator
Region	Sample Frame	Region
eapopn	Sample Frame	Sample Frame: total population in PSU
m1c_a	Final Data File	NUMBER A-# of eligible persons
m11_q07	Final Data File	Result code of interview
field_rst	Final Data File	Fieldwork result code
final_status	Final Data File	Final Result Code
PSU_HHs	STEP Ghana - Household Count per PSU	Number of listed households in PSU
PSU_parts	STEP Ghana - Household Count per PSU	Number of parts into which PSU was split

Table 6 provides the breakdown of the 6030 sample cases in the 201 activated PSUs by the final result of the interview attempt, i.e., by '*final_status*' code.

Table 6: Ghana - Number of Sampled Households by 'final_status' (activated PSUs)	
Final Status	# of Households
01-Refusal-Household	163
02-Unusual circumstance-Household	20
03-No knowledgeable household member found after 3 visits	5
04-Temporarily absent/unavailable for field period	355
05-No competent household member (e.g., severe illness, mental disability, etc.)	12
07-Dwelling could not be found / selected address has no household	4
08-Vacant dwelling / selected address has no household	20
09-Dwelling not habitable/destroyed/converted to commercial	9
21-Household Refusal to continue Household Module (begun,not complete)	2
22-Household module terminated - unusual circumstance	10
31-No eligible household member aged 15-64	62
32-Refusal by SP to begin modules 2-7	2
36-SP absent for entire survey period	9
38-SP learning or mental disability	1
39-SP speech impairment	1
51-Refusal by SP to continue modules 2-7	13
52-Interview termination-SP unusual circumstance	2
61-SP unable to read language of General Booklet (GB)	485
62-SP refusal to begin GB	100
63-SP unable to do GB due to unusual circumstance	19
64-SP blindness or visual impairment-unable to do GB	11
65-SP physical disability	3
71-SP looked through GB-unable to read/write language of GB	174
72-SP refused to continue GB after starting	317
73-GB interruption too long	18
74-GB terminated - unusual circumstance	37
75-Attempted GB, no answers marked	7
76-Attempted part or all of GB; <3 correct Core answers	604
81-No attempt at any Exercise Booklet items / refusal to start	185
82-Interruption too long during Exercise Booklet administration	3
83-Unusual circumstance, Exercise Booklet administration terminated	9
84-Attempt at some Exercise Booklet items, no answers marked	10
85-Attempt at all Exercise Booklet items, no answers marked	9
86-Marked some Exercise Booklet items, not all items attempted	248
87-Marked some Exercise Booklet items, all items attempted	398
88-Marked all Exercise Booklet items	335
95-Household not activated	2347
97-Household not interviewed - field procedure error	8
98-Incorrect eligible household member selected	13
Total	6030

4.2 Final Output Files from Weighting Process

The final output file from the weighting process is,

- 1) [GHA-W01-STEP\(HH\)-Final Design-BM Wts \(27-Nov-2013\)](#)
 - a) This file contains the original data variables from the Ghana data file '[STEP Ghana - Data for Weighting_S7.dta](#)' and the applicable PSU weights, household weights, person weights, benchmarked person weights, and other associated variables that were created during the weighting process.
 - b) The file contains a data record for each of the 3653 sampled households that were activated or should have been activated during the data collection operation.
 - i) 3013 participating households with a final household weight
 - ♦ For participating households, the weighted sum of the household weights is 2,538,956.
 - ii) 2987 participating selected persons with a final person weight
 - ♦ For participating individuals, the weighted sum of the person weights is 5,822,780.
 - iii) 2987 participating selected persons with a final benchmarked person weight
 - ♦ For participating individuals, the weighted sum of the benchmarked person weights is 7,695,441.
 - c) In addition to the Ghana 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 a list of variables from the weighting process that are included in the final weighted data file.
 - ii) Table 8 provides a list of variables from the weighting process that are not included in the final weighted data file. These variables are available from other interim files that were created during the weighting process.
 - iii) Table 9 provides the frequency count of result codes for sampled households in the final weighted data file '[GHA-W01-STEP\(HH\)-Final Design-BM Wts \(27-Nov-2013\)](#)' that were activated or should have been activated.

Table 7: Variables Created During Weighting Process - Added to Ghana Data File

Variable Name	Description	Values
m1c_a_mean	Mean # Eligible Persons in PSU	n/a
EligPers_Cnt	Eligible Person Count in Participant HH	n/a
RespCat_HH	Final Response Category-Households	1-Participant HH 2-Ineligible HH 3-Non-participant Eligible HH 4-Non-participant Unknown Eligibility HH
RespCat_SP	Final Response Category-Selected Person	1-Participant SP 2-Non-participant SP 3-Not applicable - non-participant HH
W2_FinPSUwt	Final PSU/'PSU Partition' Weight	n/a
W3_FinHHwt	Final Household Weight	n/a
W4_FinSPwt	Final Person Weight	n/a
BM2010_HHs	Total Urban Households: Census 2010	n/a
HHs_Est	STEP Estimated Total # of Households	n/a
BM_HHwt	BM Household Weight: Adjusted for Benchmark Total Households	n/a
AgeGrp	Age Group (5 years)	1 - '15 to 19' 2 - '20 to 24' 3 - '25 to 29' 4 - '30 to 34' 5 - '35 to 39' 6 - '40 to 44' 7 - '45 to 49' 8 - '50 to 54' 9 - '55 to 59' 10 - '60 to 64'
Sex_AgeGrp_Est	Persons: Estimated Sex_AgeGrp Population	n/a
Sex_AgeGrp_2010	Census Sex_AgeGrp Population(2010)	n/a
SP_BMwt	BM PERSON Weight: Adjusted for Benchmark Gender_AgeGrp Population (2010)	n/a

Table 8: Variables Created During Weighting Process – Not Added to Ghana Data File

Variable Name	Description	Values
W1_PSU	Basic PSU Weight	n/a
RespCat_PSU	Final Response Category-PSUs	n/a
W2_PSU_Partwt	PSU-"PSU Partition" Weight	n/a
W3_HH	Household Weight within PSU	n/a
W3_HHwt_basic	Basic Household Weight	n/a
W2_HH_BM	Basic HH Wt, Benchmarked to Frame # of HHs	n/a
PSU_RespFreq	Count of 'RespCat_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
Elig_Prop_HH	Proportion of Eligible HHs - known eligibility	n/a
R_adj_HH	Response adjustment ratio for HHs	n/a
EligPers_Cnt	Eligible Person Count in Sampled HH	n/a
W4_SP	Selected Person n weight within household	n/a
R_adj_SP	Response adjustment for selected person	n/a
EP_PSU_RCat	Total Eligible Persons-PSU Participant Households	n/a
EP_k	PSU - # of Participant SPs in participant HHs	n/a
EP_z	PSU - # of Non-participant SPs in participant HHs	n/a
W4_SPwt_BMHH	Final Person Weight adjusted for BM HHs	n/a

Table 9: Ghana, Frequency/Percent of Final Status Codes For [Activated](#) Sample Households

Case Final Result Code	Frequency	Percent
01-Refusal-Household	163	4.5%
02-Unusual circumstance-Household	20	0.5%
03-No knowledgeable household member found after 3 visits	5	0.1%
04-Temporarily absent/unavailable for field period	355	9.7%
05-No competent household member (e.g., severe illness, mental disability, etc.)	12	0.3%
07-Dwelling could not be found / selected address has no household	4	0.1%
08-Vacant dwelling / selected address has no household	20	0.5%
09-Dwelling not habitable/destroyed/converted to commercial	9	0.2%
21-Household Refusal to continue Household Module (begun,not complete)	2	0.1%
22-Household module terminated - unusual circumstance	10	0.3%
31-No eligible household member aged 15-64	32	0.9%
32-Refusal by SP to begin modules 2-7	2	0.1%
36-SP absent for entire survey period	9	0.2%
38-SP learning or mental disability	1	0.0%
39-SP speech impairment	1	0.0%
51-Refusal by SP to continue modules 2-7	13	0.4%
52-Interview termination-SP unusual circumstance	2	0.1%
61-SP unable to read language of General Booklet (GB)	485	13.3%
62-SP refusal to begin GB	100	2.7%
63-SP unable to do GB due to unusual circumstance	19	0.5%
64-SP blindness or visual impairment-unable to do GB	11	0.3%
65-SP physical disability	3	0.1%
71-SP looked through GB-unable to read/write language of GB	174	4.8%
72-SP refused to continue GB after starting	317	8.7%
73-GB interruption too long	18	0.5%
74-GB terminated - unusual circumstance	37	1.0%
75-Attempted GB, no answers marked	7	0.2%
76-Attempted part or all of GB; <3 correct Core answers	604	16.5%
81-No attempt at any Exercise Booklet items / refusal to start	185	5.1%
82-Interruption too long during Exercise Booklet administration	3	0.1%
83-Unusual circumstance, Exercise Booklet administration terminated	9	0.2%
84-Attempt at some Exercise Booklet items, no answers marked	10	0.3%
85-Attempt at all Exercise Booklet items, no answers marked	9	0.2%
86-Marked some Exercise Booklet items, not all items attempted	248	6.8%
87-Marked some Exercise Booklet items, all items attempted	398	10.9%
88-Marked all Exercise Booklet items	335	9.2%
97-Household not interviewed - field procedure error	8	0.2%
98-Incorrect eligible household member selected	13	0.4%
Total	3653	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 'W00-GHA Final Revised Data File(20-Nov-2013)'.

5.1 First-stage Weight (Basic PSU Weight)

The basic first stage weight, i.e., the PSU weight, is based on a sample size of 250 PSUs.

For the Ghana systematic sample of PSUs, the probability of selection of the j^{th} PSU, say p_j , may be represented as,

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

where,

y - the number of sampled PSUs (i.e., $m=250$ PSUs in the case of the Ghana STEP Survey),

m_j - the measure of size for the j^{th} PSU, in the case of the Ghana STEP Survey the measure of size for the first stage selection of PSUs was the sample frame urban population size in each PSU in the target population,

$M = 12,545,229$ is the sample frame count of the Ghana *urban* population; M is represented by

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

The basic first stage weight for the j^{th} PSU, $W1_j$, is calculated as the inverse of its probability of selection:

$$3) \quad W1_j = \frac{1}{p_j} = \frac{M}{y \cdot m_j}$$

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

Example 1: First Stage Weight

The basic first stage weight for the PSU # 139 with a population of 1510 people is calculated as follows:

$$W1_{139} = \frac{1}{p_{139}} = \frac{M}{y * m_{139}} = \frac{12,545,229}{250 * 1510} \approx 33.23$$

The basic first stage weight is similarly calculated for each of the 250 sampled PSUs.

5.1.1 First-stage Response Adjustment

The first-stage response adjustment was carried out for each sampled PSU.

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

N.B., there were no non-participating eligible PSUs in the Ghana STEP Survey.

Participant - A participating PSU is an eligible sampled PSU that provides sufficient data to be considered a participant in the survey. The sampled PSUs with one or more person-level result codes 32-40, 51-88 and 98-99 satisfy the requirement for classification of a household as a survey participant.

Ineligible – A sampled PSU is categorized as ineligible if there are no eligible household members in the PSU. PSU # 234 included only military barracks with no eligible members of the STEP target population; it was replaced by one of the 50 ‘reserve’ PSUs.

Non-participant - Eligibility Unknown: A non-participating PSU resulted when a sampled PSU was not activated. This situation pertains to the reserve sample cases; although 250 PSUs were initially selected to allow for 50 reserve PSUs for non-responding PSUs, only one reserve PSU was actually activated during data collection, thus there were 49 non-activated sampled PSUs.

The total number of sampled PSUs may be represented by the following equation.

$$4) \quad y_j = y_k + y_x + y_u$$

where,

y_j - the number of sampled PSUs, (i.e., $y_j = 250$ PSUs in the Ghana STEP),

y_k - the number of participant PSUs ‘k’,

y_x - the number of ineligible PSUs ‘x’,

y_u - the number of non-participating PSUs ‘u’ with unknown eligibility.

Table 10 provides the STEP count of sampled PSUs from the Ghana data file by PSU Response Category.

Table 10: Ghana Number of PSUs by PSU Response Category	
PSU Response Category	# of PSUs
1-Participant PSU (y_k)	200
2-Ineligible PSU (y_x)	1
3-Non-participant PSU; Unknown Eligibility (y_u)	49
Total (y_j)	250

5.1.1.1 Eligible Proportion of PSUs

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

Amongst the y_u non-participating PSUs with unknown eligibility, the proportion of PSUs with known eligibility is estimated to be

$$5) \quad \varepsilon = \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} = \frac{y_k}{y_k + y_x} = \frac{200}{200 + 1} = 0.995$$

- Note that the eligible proportion is the same for all eligible participating PSUs.

5.1.1.2 Response Adjustment Factor for Non-participation

A PSU response adjustment is necessary to compensate for PSUs that did not participate in the survey. The PSU response adjustment factor for the j^{th} PSU, R_j , is calculated as follows:

$$6) \quad R_{PSU_j} = \frac{\text{Total eligible PSUs}}{\text{Total participating PSUs}} = \frac{y_k + (\varepsilon * y_u)}{y_k} = \frac{200 + (0.995 * 49)}{200} \approx 1.24378$$

- The response adjustment ratio, R_{PSU_j} , incorporates the adjustment for unknown eligibility, ε .
➤ The PSUs that are categorized as ineligible, y_x , are excluded from the calculation of the PSU response adjustment.

5.1.2 Response-adjusted PSU Weight

The response-adjusted PSU weight is calculated as the product of the basic PSU weight and the PSU response adjustment factor. Therefore, for PSU 'j', the response-adjusted PSU, W_{PSU_j} , is calculated as follows:

$$7) \quad W_{1R_j} = W_{1j} * R_{PSU_j}$$

Example 2: Response-adjusted PSU Weight

For PSU #139,

$$\begin{aligned} W_{1R_j} &= W_{1j} * R_{PSU_j} \\ &\approx 33.23 \times 1.24378 \\ &\approx 41.33 \end{aligned}$$

5.2 Second-stage Weight (PSU Partition Weight)

The second stage unit is a PSU partition. The 201 PSUs were divided into one to three PSU partitions, and one PSU partition was selected from each set of PSU partitions.

Note: The second stage weight is calculated under the assumption that each sampled PSU was divided into one to three PSU partitions with the same urban population size in each PSU partition, and one partition was selected from each sampled PSU.

The basic second stage weight, $W2_{js}$, for the s^{th} PSU partition in PSU 'j' is calculated as follows:

$$8) \quad W2_{js} = \frac{s_j}{s_j}$$

where,

s_j - the number of PSU partitions in the j^{th} PSU,

s_j - the number of sampled PSU partitions from the j^{th} PSU (i.e., $s_j = 1$ for all activated PSUs).

Example 3: Second Stage PSU Partition Weight

The basic second stage weight for the *urban* PSU # 139 which was divided into 2 PSU partitions is calculated as follows:

$$W2_{js} = \frac{1}{p_{js}} = \frac{s_j}{s_j} = \frac{2}{1} = 2$$

The PSU Partition weight is similarly calculated for all sampled PSU Partitions.

5.2.1 Final PSU Weight

The final PSU weight is calculated as the product of the response-adjusted PSU weight and the PSU Partition weight. Therefore, for PSU 'j', the final PSU/PSU Partition weight, $WPSU_{js}$, is calculated as follows:

$$9) \quad WPSU_{js} = W1_{R_j} * W2_{js}$$

➤ The final PSU/PSU Partition weight is appended to the STEP data file.

Example 4: Final PSU Weight

For PSU #139,

$$\begin{aligned} WPSU_{js} &= W1_{R_j} * W2_{js} \\ &\approx 41.33 \times 2 \\ &\approx 82.67 \end{aligned}$$

The final PSU weight is similarly calculated for all sampled PSUs.

5.3 Third-stage Weight (Household Weight)

5.3.1 Basic Household Weight

The basic third stage weight for the k^{th} household in PSU ' j ', $W3_{jk}$, is calculated as follows:

$$10) \quad W3_{jk} = \frac{N_j}{n_j}$$

where,

N_j - the number of listed households in the j^{th} PSU,

n_j - the number of households sampled in the j^{th} PSU, (i.e., $n_j = 30$ households in each Ghana sampled PSU).

Example 5: Third Stage Basic Household Weight

The basic third stage weight for each of the 30 sampled households in the Ghana PSU # 139 which contains 109 listed households is calculated as follows:

$$W3_{jk} = \frac{N_j}{n_j} = \frac{109}{30} \approx 4.63$$

The basic third stage weight is similarly calculated for each of the PSUs/PSU partitions.

5.3.2 Third-stage Response Adjustment

The third-stage response adjustment was carried out for each participant PSU/PSU partition.

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 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 and 98-99 satisfy the requirement for classification of a 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
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
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
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 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
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.

$$11) \quad n_j = n_{jk} + n_{jx} + n_{jz} + n_{ju}$$

where,

n_j - the number of households sampled in the j^{th} PSU, (i.e., $n_j = 30$ households in each Ghana sampled PSU),

n_{jk} - the number of households ‘k’ that participate in the j^{th} PSU,

n_{jx} - the number of ineligible households ‘x’ in the j^{th} PSU,

n_{jz} - the number of non-participating households ‘z’ with known eligibility in the j^{th} PSU,

n_{ju} - the number of non-participating households ‘u’ with unknown eligibility in the j^{th} PSU.

Table 11 provides the STEP counts of sampled households from the Ghana data file by the outcome categories described above.

Table 11: Number of Sampled Households by Final Response Category						
Case Final Result Code	Final Response Category-Households					% of Total
	Participant (n _{jk})	Ineligible (n _{jk})	Non-participant Eligible (n _{jz})	Non-participant Eligibility Unknown (n _{ju})	Total (n _j)	
01-Refusal-Household				163	163	2.7%
02-Unusual circumstance-HH				20	20	0.3%
03-No knowledgeable HH member found				5	5	0.1%
04-Temporarily absent				355	355	5.9%
05-No competent HH member				12	12	0.2%
07- Selected address has no household				4	4	0.1%
08-Vacant dwelling		20			20	0.3%
09-Dwelling not habitable		9			9	0.1%
21-HH Refusal to continue HH Module			2		2	0.0%
22-HH module terminated - unusual circumstance			10		10	0.2%
31-No eligible HH member aged 15-64		32			32	0.5%
32-Refusal by SP to begin modules 2-7	2				2	0.0%
36-SP absent for entire survey period	9				9	0.1%
38-SP learning or mental disability	1				1	0.0%
39-SP speech impairment	1				1	0.0%
51-Refusal by SP to continue modules 2-7	13				13	0.2%
52-Interview termination-SP unusual circumstance	2				2	0.0%
61-SP unable to read General Booklet (GB) language	485				485	8.0%
62-SP refusal to begin GB	100				100	1.7%
63-SP unable to do GB due to unusual circumstance	19				19	0.3%
64-SP visual impairment-unable to do GB	11				11	0.2%
65-SP physical disability	3				3	0.0%
71-SP unable to read/write language of GB	174				174	2.9%
72-SP refused to continue GB	317				317	5.3%
73-GB interruption too long	18				18	0.3%
74-GB terminated - unusual circumstance	37				37	0.6%
75-Attempted GB, no answers marked	7				7	0.1%
76-Attempted GB; <3 correct Core answers	604				604	10.0%
81-No attempt at Exercise Booklet(EB) items	185				185	3.1%
82-Interruption too long-EB termination	3				3	0.0%
83-Unusual circumstance, EB terminated	9				9	0.1%
84-Attempt at some EB items, no answers marked	10				10	0.2%
85-Attempt at all EB items, no answers marked	9				9	0.1%
86-Marked some EB items, some items attempted	248				248	4.1%
87-Marked some EB items, all items attempted	398				398	6.6%
88-Marked all EB items	335				335	5.6%
95-Household not activated				2347	2347	38.9%
97-HH not interviewed - field procedure error			8		8	0.1%
98-Incorrect eligible HH member selected	13				13	0.2%
Total	3013	61	20	2906	6000	100.0%

5.3.2.1 Eligible Proportion of Households

The STEP Survey assumes that the households 'u' with unknown eligibility in the j^{th} PSU are comprised of some proportion, say ε_j , that are eligible households and the complementary proportion, $1 - \varepsilon_j$, that are ineligible households.

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

$$12) \quad \varepsilon_j = \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} = \frac{n_{jk} + n_{jz}}{n_{jk} + n_{jx} + n_{jz}}$$

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

Example 6: Eligible Proportion of Households

For PSU #139, the following household results were obtained:

n_j	n_{jk}	n_{jx}	n_{jz}	n_{ju}
30	15	1	1	13

For PSU # 139, the proportion of households with known eligibility is calculated as,

$$\begin{aligned} \varepsilon_{139} &= \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} \\ &= \frac{n_{jk} + n_{jz}}{n_{jk} + n_{jx} + n_{jz}} \\ &= \frac{15 + 1}{15 + 1 + 1} = \frac{16}{17} \approx 0.941176 \end{aligned}$$

5.3.2.2 Response Adjustment Factor 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 factor for the j^{th} PSU, R_j , is calculated as follows:

$$13) \quad R_{HHj} = \frac{\text{Total eligible households in PSU } j}{\text{Total participating households in PSU } j} = \frac{n_{jk} + n_{jz} + (\epsilon_j * n_{ju})}{n_{jk}}$$

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

Example 7: Second Stage Response Adjustment Factor for sampled Households

For example, the following results were obtained for PSU #139

n_j	n_{jk}	n_{jx}	n_{jz}	n_{ju}
30	15	1	1	13

Therefore,

$$\begin{aligned}
 R_{102} &= \frac{\text{Total eligible households in PSU \#139}}{\text{Total participating households in PSU \#139}} \\
 &= \frac{n_{jk} + n_{jz} + (\epsilon_{102} * n_{ju})}{n_{jk}} \\
 &= \frac{15 + 1 + (.941176 * 13)}{15} \approx 1.8824
 \end{aligned}$$

The response adjustment ratio, R_{HHj} , is similarly calculated for all sampled PSUs.

5.3.3 Final Third-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 third stage household weight is calculated as the product of the final PSU-‘PSU Partition’ weight, the basic household weight, and the response adjustment factor for non-participant households. Therefore, for household ‘k’ in PSU ‘j’, the final third stage household weight, $W3_{HH_k}$, (i.e., weighted data file variable **W3_FinHHwt**) is calculated as follows:

$$14) \quad W3_{HH_k} = WPSU_{js} * W3_{jk} * R_{HH_j}$$

➤ The final third stage household weight is appended to the STEP data file.

Example 8: Third Stage Response-adjusted Household Weight

For PSU #139,

$$\begin{aligned} W3_{HH_k} &= WPSU_{js} * W3_{jk} * R_{HH_j} \\ &\approx 82.67 \times 3.63 \times 1.8824 \\ &\approx 565.38 \end{aligned}$$

The household weight is similarly calculated for all sampled households.

5.4 Fourth-stage Weight (Person Weight)

5.4.1 Basic Fourth-stage Person Weight

The basic fourth stage weight for the i^{th} person, $W4_{jki}$, is calculated as follows:

$$15) \quad W4_{jki} = \frac{a_{jk}}{b_{jk}} = \frac{a_{jk}}{1} = a_{jk}$$

where,

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

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

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

where,

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

Example 9: Fourth Stage – Basic Person Weight within sampled Household

For PSU #139, selected household # 04 has 4 eligible persons. Therefore,

$$W4_{jki} = \frac{a_{jk}}{b_{jk}} = \frac{4}{1} = 4$$

5.4.2 Fourth-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 51-88 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
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:

II. Person-level Codes – Non-participant in Eligible Selected Household	
Code	Code Description
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
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 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

Note: there were no ineligible selected persons in the Ghana STEP Survey.

The fourth stage person weight is only relevant for sampled households in the final response category RespCat_HH=1, i.e., participant households. Table 12 provides the pertinent frequency counts of the Final Status Codes for the participant households.

Table 12: Final Selected Person Status Results for Participant Households		
Selected Person Final Result Code	n	% of Total n
32-Refusal by SP to begin modules 2-7	2	.1%
36-SP absent for entire survey period	9	.3%
38-SP learning or mental disability	1	.0%
39-SP speech impairment	1	.0%
51-Refusal by SP to continue modules 2-7	13	.4%
52-Interview termination-SP unusual circumstance	2	.1%
61-SP unable to read language of General Booklet (GB)	485	16.1%
62-SP refusal to begin GB	100	3.3%
63-SP unable to do GB due to unusual circumstance	19	.6%
64-SP blindness or visual impairment-unable to do GB	11	.4%
65-SP physical disability	3	.1%
71-SP looked through GB-unable to read/write language of GB	174	5.8%
72-SP refused to continue GB after starting	317	10.5%
73-GB interruption too long	18	.6%
74-GB terminated - unusual circumstance	37	1.2%
75-Attempted GB, no answers marked	7	.2%
76-Attempted part or all of GB; <3 correct Core answers	604	20.0%
81-No attempt at any Exercise Booklet items / refusal to start	185	6.1%
82-Interruption too long during Exercise Booklet administration	3	.1%
83-Unusual circumstance, Exercise Booklet administration terminated	9	.3%
84-Attempt at some Exercise Booklet items, no answers marked	10	.3%
85-Attempt at all Exercise Booklet items, no answers marked	9	.3%
86-Marked some Exercise Booklet items, not all items attempted	248	8.2%
87-Marked some Exercise Booklet items, all items attempted	398	13.2%
88-Marked all Exercise Booklet items	335	11.1%
98-Incorrect eligible household member selected	13	.4%
Total	3013	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, $W4_{jki}$, 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:

$$16) \quad A_{jk} = A_{jki} + A_{jkq}$$

where,

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

$$A_{jki} = \sum_{k=1}^{n_{jk}} a_{jki} \quad \text{- the total number of participant eligible persons 'i' in the } j^{\text{th}} \text{ PSU,}$$

$$A_{jkq} = \sum_{k=1}^{n_{jk}} a_{jkq} \quad \text{- the total number of non-participating eligible persons 'q' in the } j^{\text{th}} \text{ PSU,}$$

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

$$17) \quad R_{SP_j} = \frac{\text{Total eligible persons in all participating sampled households in PSU 'j'}}{\text{Total participant persons in all participating sampled households in PSU 'j'}} = \frac{A_{jk}}{A_{jki}}$$

Example 10: Third Stage – Basic Person Weight Response Adjustment

- i) For PSU #139, there are 34 eligible persons across all participating households; also, all selected persons participated. Therefore,

$$A_{jk} = \sum_{k=1}^{n_{jk}} a_{jk} = 34 ; \quad A_{jki} = \sum_{k=1}^{n_{jk}} a_{jki} = 34$$

$$R_{SP_j} = \frac{A_{jk}}{A_{jki}} = \frac{34}{34} = 1.00$$

- ii) For PSU # 259, the total number of eligible persons across all participating households is 48. For PSU # 259, there is one non-participating selected persons Therefore,

$$A_{jk} = \sum_{k=1}^{n_{jk}} a_{jk} = 48 ; \quad A_{jki} = \sum_{k=1}^{n_{jk}} a_{jki} = 43$$

$$R_{SP_j} = \frac{A_{jk}}{A_{jki}} = \frac{48}{43} \approx 1.116$$

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 HHs		
For Participating Households		
Final Response Category-SP	Frequency	Percent
1-Participant SP	2987	99.1
2-Non-participant, Eligible	26	0.9
Total	3013	100.0

Based on the above table, there are 2987 participating selected persons for which a final non-zero

Person-weight was calculated. The sampled persons in category 2 have a zero-value final person weight.

5.4.3 Final Fourth-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 fourth 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:

$$18) \quad W_{4SP} = W_{3HH_k} * W_{4jki} * R_{SP_j}$$

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

Example 11: Third Stage – Final Person Weight

- i) For PSU #139, selected household # 04 with 3 eligible persons, the selected person was a participant. All selected persons participated in PSU #139, therefore, $R_{SP_j} = 1$

$$W_{4SP} = 565.3818 * 4 * 1 \approx 2261.5275$$

- ii) For PSU #259, in participant household # 24 with 6 eligible persons, the selected person was a participant. For this household,

$$W_{3HH_k} \approx 492.8158; W_{4SP} \approx 6; \quad R_{SP_j} \approx 1.1162$$

$$W_{4SP} = 492.8158 * 6 * 1.1162 \approx 3300.72$$

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.

6 Benchmark Weights – Adjustment of Weights to Known Population Totals

The STEP weights were adjusted for known population totals from the Ghana Census 2010. The STEP final household weights were adjusted to the known total number of urban households obtained by the 2010 Census. The STEP final person weights were adjusted to the known Census 2010 urban population by gender and 5 year age groups for persons aged 15 to 19 years of age.

Note: There may be reasonable explanations for the differences between the STEP Survey and the Census 2010 with regard to the number of urban households and the urban population aged 15 to 19.

For example, the differences may be attributable to

- 1) Difference in Household Definition: Census definition of a household member is not exactly the same as the STEP definition of a household;
- 2) Frame timeliness (i.e., up-to-date issue), for example,
 - a) The STEP sample was selected in 2012, two years after the 2010 Census; there could have been differences in the number of PSU listed households for Step versus the number from the 2010 Census due to changes during the 2 year period since the 2010 Census.
 - b) Census process for classification of EAs as urban or rural may have resulted in changes to the urban PSU counts obtained by the STEP field staff.

6.1 Benchmark-adjusted Household Weight

The product of the basic third stage household weight and the final PSU weight provides the estimated number of households, \hat{N}_{HH} , in the target population, i.e.,

$$\hat{N}_{HH} = \sum W3_{HH_k} \cdot f \text{ for the Ghana sample of 201 participating PSUs,}$$

$$\hat{N}_{HH} = 2,538,955.596 \text{ households.}$$

According to the 2010 Ghana Census, the number of households in the Ghana urban target population is $N_{HH} = 3,049,366$ households.

In order to make \hat{N} consistent with the 2010 Ghana Census count of households a benchmark-adjusted household weight, $W3_{BM_j}$, was calculated for each PSU 'j' as follows:

$$19) \quad W3_{BM_j} = W3_{HH_k} * \frac{N_{HH}}{\hat{N}_{HH}}$$

Example 12: Benchmark-adjusted Household Weight

For PSU #139,

$$\begin{aligned} W3_{BM_j} &= W3_{HH_k} * \frac{N_{HH}}{\hat{N}_{HH}} \\ &= 565.3819 \times \frac{3,049,366}{2,538,955.596} \\ &\approx 679.0415 \end{aligned}$$

The benchmark adjusted household weight is similarly calculated for all sampled households.

6.2 Benchmark-adjusted Person Weight

6.2.1 Person Weight Adjustment for Known Number of Households

Firstly, the benchmark adjustment of the STEP final person weights included an adjustment that is based on the benchmark-adjusted household weight.

$$20) \quad W4_{SP_BMHH} = W4_{SP} * \frac{W3_{BMj}}{W3_{HH_k}}$$

where,

$W4_{SP_BMHH}$ - the final STEP person weight adjusted for the known Census 2010 number of households;

$W4_{SP}$ - the final STEP person weight, unadjusted for known Census 2010 population totals;

$W3_{BMj}$ - the STEP household weight adjusted for the 2010 Census household count in the target population;

$W3_{HH_k}$ - the STEP household weight, unadjusted for known Census 2010 population totals.

Example 13: Person Weight, Adjusted for the Known Number of Households

For PSU #139, household #11 has 3 eligible persons; the selected person, a 54 year old male, is a participant. Thus,

$$\begin{aligned} W4_{SP_BMHH} &= W4_{SP} * \frac{W3_{BMj}}{W3_{HH_k}} \\ &= 1696.1456 \times \frac{679.0415}{565.3819} \\ &\approx 2037.1246 \end{aligned}$$

6.2.2 Person Weight: Adjustment for Known Number of Persons, Gender, Age Group

Secondly, for each Gender and Age Group (i.e., GA) combination, the Ghana weighted data file includes the variable 'Sex_AgeGrp_Est', the STEP Survey estimate, say \hat{Y}_{GA} , of the target population in each GA combination. Note that \hat{Y}_{GA} is based on $W3_{BM_j}$, the STEP household weight adjusted for the 2010 Census household count in the target population.

Table 14 provides the weighted estimate of the number of persons in the STEP target population. The table includes (i) the Step Estimated Population, based on the STEP final person weight (unadjusted for the known number of Census 2010 household total); (ii) the Step Estimated Population, based on the person weight, adjusted for the known number of Census 2010 household total.

Table 14: STEP Estimates of Urban Population by Gender & Age Group (Unadjusted & Adjusted for Known Number of Households)

Age Group	# of Sample Cases			STEP Estimated Population			STEP Estimated Population		
				Based on STEP Final Person Weight			Based on Person Weight, adjusted for Census 2010 # of HHs (Sex_AgeGrp_Est ¹)		
(5 years)	Male	Female	Total	Male	Female	Total	Male (\hat{Y}_{1A})	Female (\hat{Y}_{2A})	Total
01 - '15 to 19'	164	185	349	414,631	423,392	838,023	497,986	508,507	1,006,492
02 - '20 to 24'	201	262	463	430,985	604,342	1,035,327	517,627	725,833	1,243,460
03 - '25 to 29'	247	306	553	356,563	498,246	854,809	428,244	598,409	1,026,652
04 - '30 to 34'	199	272	471	293,805	492,707	786,512	352,869	591,757	944,625
05 - '35 to 39'	135	231	366	252,057	417,589	669,646	302,729	501,538	804,266
06 - '40 to 44'	91	146	237	161,178	282,301	443,479	193,580	339,052	532,632
07 - '45 to 49'	69	100	169	127,497	234,425	361,922	153,128	281,552	434,680
08 - '50 to 54'	61	67	128	138,403	142,323	280,726	166,227	170,934	337,161
09 - '55 to 59'	66	70	136	147,513	121,164	268,677	177,168	145,521	322,689
10 - '60 to 64'	47	68	115	129,827	153,832	283,659	155,926	184,757	340,683
Total	1280	1707	2987	2,452,460	3,370,320	5,822,780	2,945,482	4,047,861	6,993,342

1 - 'Sex_AgeGrp_Est' is based on the STEP household weight adjusted for the known # of households in the 2010 Census

The Ghana Census 2010 publication '*2010 Population & Housing Census: Demographic, Social, Economic & Housing Characteristics*' includes the Ghana Census population by gender and age for the year 2010. This publication was used to generate, the variable 'Sex_AgeGrp_2010', say Y_{GA} , the census urban population for each GA combination corresponding to the variable 'Sex_AgeGrp_Est'.

In order to make \hat{Y}_{GA} consistent with the Ghana Census population count, a final benchmark-adjusted person weight, $W4_{SP_{BM}}$, was calculated for each PSU 'j' as follows:

$$21) \quad W4_{SP_{BM}} = W4_{SP_{BMHH}} * \frac{Y_{GA}}{\hat{Y}_{GA}}$$

Table 15 provides the data that was used to adjust the person weight to Ghana Census urban population figures for the STEP target population; it includes, for each GA, the Step estimated urban population and the corresponding Census urban population.

Table 15: STEP Estimated Urban Population vs Ghana Census Urban Population (2010) by Gender & Age Group

Age Group (5 years)	# of Sample Cases			STEP <u>Estimated</u> Urban Population			Ghana Census Urban Population		
				Sex_AgeGrp_Est ¹			Sex_AgeGrp_2010		
	Male	Female	Total	Male \hat{Y}_{1A}	Female \hat{Y}_{2A}	Total	Male Y_{1A}	Female Y_{2A}	Total
01 - '15 to 19'	164	185	349	497,986	508,507	1,006,492	651,829	712,295	1,364,124
02 - '20 to 24'	201	262	463	517,627	725,833	1,243,460	642,140	714,698	1,356,838
03 - '25 to 29'	247	306	553	428,244	598,409	1,026,652	553,927	640,940	1,194,867
04 - '30 to 34'	199	272	471	352,869	591,757	944,625	455,204	500,494	955,698
05 - '35 to 39'	135	231	366	302,729	501,538	804,266	373,678	408,174	781,852
06 - '40 to 44'	91	146	237	193,580	339,052	532,632	304,875	327,174	632,049
07 - '45 to 49'	69	100	169	153,128	281,552	434,680	231,146	256,046	487,192
08 - '50 to 54'	61	67	128	166,227	170,934	337,161	195,556	225,922	421,478
09 - '55 to 59'	66	70	136	177,168	145,521	322,689	132,984	142,974	275,958
10 - '60 to 64'	47	68	115	155,926	184,757	340,683	105,738	119,647	225,385
Total	1280	1707	2987	2,945,482	4,047,861	6,993,342	3,647,077	4,048,364	7,695,441

1 - 'Sex_AgeGrp_Est' is based on the STEP household weight adjusted for the known # of households in the 2010 Census

Example 14: Final Benchmark-adjusted Person Weight

For PSU #139, the participant selected person in selected household # 11 is a 54 year old male. Thus,

$$\begin{aligned}
 W4_{SP_{BM}} &= W4_{SP_{BMHH}} * \frac{Y_{GA}}{\hat{Y}_{GA}} \\
 &= 2037.1246 \times \frac{195556}{166226.9214} \\
 &\approx 2396.5548
 \end{aligned}$$

7 Weighted Estimates from STEP

7.1 Weighted Household Count

The original sample frame 'Frame_Aug' showed there is a total urban population of 12,545,229 in Ghana. According to information provided by the survey firm, the average urban household size is 4.4 persons per household; thus, the estimated number of urban households in Ghana is

$$\text{Estimated \# of HHs in Ghana Sample Frame} = \frac{12,545,229}{4.4} = 2,851,188 \text{ households}$$

However, the STEP data collection found that some of the target population households are ineligible for reasons such as vacant, not habitable, no eligible household member, etc.

Table 16 provides,

- the weighted count of households based on the sample frame count of households for the 200 participant PSUs;
- the weighted count of households based on the final status of the cases in the Ghana data file.

Table 16: Ghana – Weighted Number of Households – Sample Frame vs. STEP Data File		
Source of PSU Household Count	# of PSUs	Estimated # of Households
1. Based on # PSU of Hhlds in Sample Frame 'Frame_Aug'	16,503	2,851,188
2. Based on the household weight for all sampled households [all households, i.e., participants, non-participants, ineligible]	200	2,589,556
3. Based on the final household weight for participant households [eligible households only]	200	2,538,956
4. Based on the final benchmark-adjusted household weight for participant households [eligible households only]	200	3,049,366

As shown in the third row of Table 16, when ineligible households are excluded, the weighted number of eligible households is **2,538,956**. The fourth row of the table provides the weighted estimate of the number of households based on the final bench-mark adjusted household weight.

For participating selected households, Table 17 provides a summary of the central tendency for the participating households.

Table 17: Summary - Household Weights		
Statistic	Final Household Weight	Final Benchmark Household Weight
	(No benchmark adjustment)	(Adjusted for known # of Households)
No. of Participant Households	3013	3,013
Mean	842.67	1,012.07
Std. Error of Mean	7.48	8.99
Median	774.31	929.98
Minimum	153.94	184.88
Maximum	3327.68	3,996.65
Sum of Weights	2,538,956	3,049,366.00

7.2 Weighted Person Count

The weighted count of the non-institutionalized persons 15 to 64 years of age (inclusive) living in private dwellings in urban areas of **Ghana** at the time of data collection is provided in Table 18. Based on the 2987 participating selected persons, the sum of the final person weights from the STEP survey is 5,822,780; the sum of the benchmark-adjusted final person weights from the STEP survey is 7,695,441.

Table 18: Sample Frequency & Weighted Persons Totals by Final Selected Person Response Category For Participating Households			
Response Category (SP)	Sampled Cases	Total # of Eligible Persons in Target Population	
		(based on Final Person weight)	(based on Benchmark-adjusted Person weight)
1-Participant SP	2,987	5,822,780	7,695,441
2-Non-participant, Eligible	26	0	0
Total	3,013	5,822,780	7,695,441

For participating selected persons, Table 19 provides a summary of the central tendency of the number of eligible persons in the participating households.

Table 19: Summary - Person Weights, Eligible Persons				
Statistic	Eligible Persons in Participant HHs	Final Person Weight	Final Person Weight adjusted for BM HHs	Final Benchmark Person Weight
		(No benchmark adjustment)	(Adjusted for known # of HHs)	(Adjusted for known # of HHs <u>and</u> known population by Gender & Age Group)
No. of Respondents	2987	2987	2987	2987
Mean	2.3418	1949.37	2341.26	2576.31
Std. Error of Mean	.02648	28.75	34.53	38.16
Median	2.0000	1523.72	1830.04	1992.15
Minimum Weight	1.00	153.94	184.88	156.37
Maximum Weight	15.00	19966.09	23979.91	24634.52
Sum of Weights	6995	5,822,780	6,993,342	7,695,441

8 Response Rate

An overall response rate of 83.2% was achieved in the Ghana STEP Survey. Table 20 provides the detailed percentage distribution by final status code.

Table 20: Response Rate: Ghana, Activated HHs by Final Status Code

Case Final Status Code	Frequency	Percent
07-Dwelling could not be found / selected address has no household	4	0.1%
08-Vacant dwelling / selected address has no household	20	0.5%
09-Dwelling not habitable/destroyed/converted to commercial	9	0.2%
31-No eligible household member aged 15-64	32	0.9%
Ineligible HHs (N.B. % is based on total # of activated HHs, i.e., 3653 HHs)	65	1.8%
01-Refusal-Household	163	4.5%
02-Unusual circumstance-Household	20	0.6%
03-No knowledgeable household member found after 3 visits	5	0.1%
04-Temporarily absent/unavailable for field period	355	9.9%
05-No competent household member (e.g., severe illness, mental disability, etc.)	12	0.3%
21-Household Refusal to continue Household Module (begun, not complete)	2	0.1%
22-Household module terminated - unusual circumstance	10	0.3%
32-Refusal by SP to begin modules 2-7	2	0.1%
36-SP absent for entire survey period	9	0.3%
38-SP learning or mental disability	1	0.0%
39-SP speech impairment	1	0.0%
97-Household not interviewed - field procedure error	8	0.2%
98-Incorrect eligible household member selected	13	0.4%
Non-respondents (Eligible) (N.B. % is based on total # of Eligible HHs, i.e., 3588 HHs)	601	16.8%
51-Refusal by SP to continue modules 2-7	13	0.4%
52-Interview termination-SP unusual circumstance	2	0.1%
61-SP unable to read language of General Booklet (GB)	485	13.5%
62-SP refusal to begin GB	100	2.8%
63-SP unable to do GB due to unusual circumstance	19	0.5%
64-SP blindness or visual impairment-unable to do GB	11	0.3%
65-SP physical disability	3	0.1%
71-SP looked through GB-unable to read/write language of GB	174	4.8%
72-SP refused to continue GB after starting	317	8.8%
73-GB interruption too long	18	0.5%
74-GB terminated - unusual circumstance	37	1.0%
75-Attempted GB, no answers marked	7	0.2%
76-Attempted part or all of GB; <3 correct Core answers	604	16.8%
81-No attempt at any Exercise Booklet items / refusal to start	185	5.2%
82-Interruption too long during Exercise Booklet administration	3	0.1%
83-Unusual circumstance, Exercise Booklet administration terminated	9	0.3%
84-Attempt at some Exercise Booklet items, no answers marked	10	0.3%
85-Attempt at all Exercise Booklet items, no answers marked	9	0.3%
86-Marked some Exercise Booklet items, not all items attempted	248	6.9%
87-Marked some Exercise Booklet items, all items attempted	398	11.1%
88-Marked all Exercise Booklet items	335	9.3%
Respondents (N.B. % is based on total # of Eligible HHs, i.e., 3588 HHs)	2986	83.2%
Total Eligible Households	3588	100.0%
Total Activated Households	3653	100.0%