

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

Vietnam

October 11, 2013

Table of Contents

1	Survey Design Overview.....	1
2	Data Collection Outcomes	2
3	Weighting Overview.....	4
4	Weighting Process Files.....	5
4.1	Input Files.....	5
4.2	Revisions to Input Data File	6
4.3	Final Output Files from Weighting Process.....	8
5	Weight Calculations based on World Bank (WB) Weight Requirement.....	10
5.1	Phase 1: PSU Weight.....	10
5.2	Phase 1: Household Weight.....	10
5.3	Phase 2: PSU Weight PSU Weight.....	11
5.4	Phase 2: Basic Household Weight for Each PSU	12
5.4.1	Phase 2: Household Weight Response Adjustment.....	13
5.4.2	Phase 2: Household Weight - Adjustment for Unknown Eligibility	15
5.4.3	Phase 2: Adjustment of Household Weights for Non-participation	15
5.4.4	Household Weight - Final.....	16
5.4.5	Weighted Household Counts	17
5.5	Phase 2: Person Weight	18
5.5.1	Phase 2: Person Weight - Person-level Response Adjustment.....	19
5.5.2	Phase 2: Final Person Weight.....	22
5.5.3	Weighted Person Counts	23

October 11, 2013

1 Survey Design Overview

The Viet Nam 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 Viet Nam survey firm set a sample goal of 3405 acceptable interviews, thus exceeding the minimum sample requirement.

The target population for the Viet Nam STEP survey comprised all non-institutionalized persons 15 to 64 years of age (inclusive) living in private dwellings in urban areas of **Hanoi and Ho Chi Minh City** at the time of data collection.

The Viet Nam STEP Survey is a 2 phase sample design. The Phase 1 sample was previously carried out by the General Statistics Office (GSO) of Viet Nam to create a master sample list of 15% of EAs for all of Viet Nam. The STEP Survey selected a sample of EAs from the master list of 15% of the EAs for Hanoi and Ho Chi Minh City that was created during Phase 1. At the second phase the STEP Survey employed a 3 stage cluster sample design.

For Phase 2, the sample frame for the selection of first stage sample units was the Excel file '[VNM Sample Frame-HN-HCN-20-April-2012](#)' that was provided by the Viet Nam survey firm. The frame is a master sample list of 15% of the first stage sampling units, i.e., Enumeration Areas (EAs), in the urban areas of Hanoi and Ho Chi Minh City. The source of this sample frame is Viet Nam's 2009 Population and Housing Census.

Phase 2 First Stage Sample

The primary sample unit (PSU) is a Census Enumeration Area (EA). The sampling objective was to conduct interviews in 227 EAs. At the first stage of sample selection, 23 additional EAs were also selected as reserve PSUs to be used in the event that it was impossible to obtain any interviews in one or more of the initial PSUs. A total of 250 PSUs were selected with probability proportional to size, where the measure of size was the number of households in a PSU. Subsequently, from the file of 250 sampled PSUs, a PPS sample of 227 PSUs was selected to be the 'Initial' PSU sample.

Note: None of the 23 reserve PSUs was activated during data collection.

Phase 2 Second Stage Sample

The second stage sample unit (SSU) is a household. The sampling objective was to obtain interviews at 15 households within each selected PSU, resulting in a final initial sample of 3405 interviews. At the second stage of sample selection, 30 households were selected in each PSU using a systematic random method. The 30 households were randomly divided into 15 'Initial' households, and 15 'Reserve' households that were ranked according to the random sample selection order.

Phase 2 Third Stage Sample

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

2 Data Collection Outcomes

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

Table 1: <i>Full Literacy Assessment</i> Result Codes for Sampled Households and Selected Persons	
I. Household-level Codes	
Code	Code Description
01	Household Refusal
02	Household unusual circumstance (e.g., death in family, illness, fire in dwelling, etc.)
03	No knowledgeable household member found after 3 visits (only a child, non-competent adult, etc.)
04	Temporarily absent/unavailable for field period (information from other persons)
05	No competent household member to interview (because of severe illness, mental disability, etc.)
06	Language problem – no household member spoke a language understandable by survey team, no translator available
07	Dwelling could not be found / selected address has no household
08	Vacant dwelling
09	Dwelling not habitable/dwelling destroyed/dwelling converted to commercial use
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
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

Table 1: <i>Full Literacy Assessment</i> Result Codes for Sampled Households and Selected Persons	
65	Physical disability of Selected Person prevented completion of General Booklet
	General Booklet (module 9) looked through or attempted but not completed and not passed
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
	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 Viet Nam Phase 2 three-stage stratified cluster design greatly enhanced the operational feasibility of data collection, it resulted in differential probabilities of selection for the selected persons. Consequently, each selected person in the survey does not necessarily represent the same number of persons in the target population. To account for differential probabilities of selection due to the nature of the design and to ensure accurate survey estimates, STEP requires a sampling weight for each person that participated in the survey.

In general, the objectives of the STEP weighting are to construct a set of survey weights to,

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

The general weighting procedure for the Viet Nam 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, and an Exercise Booklet, and all household and individual questionnaire modules.

- a) Calculation of a PSU weight for 227 sampled PSUs;
 - b) Calculation of a household weight for each sampled household;
 - i) Calculation of a household-level non-response adjustment independently for each PSU.
 - c) Calculation of a person weight for each selected person (SP);
 - i) Calculation of a non-response adjustment independently for each sampled person.
- 3) The required output from the weighting process is a final Viet Nam data file with the survey design weights (i.e., for each sampled PSU, household, person) appended to each data record.

4 Weighting Process Files

4.1 Input Files

The key file for the weighting process is the STATA file '[STEP - Vietnam - Data for Weights - Revised Jan 2 2012.dta](#)' provided by the WB core team. An SPSS version of this file with a revised name, i.e., '[STEP - Vietnam - Data for Weights - Revised Jan 2 2013-1](#)', was created for the weighting process. This file contains 6810 sample cases, i.e., 3405 Initial sampled cases and 3405 Reserve samples cases. This file is the primary dataset for input to the weighting process.

Table 2 provides the frequency counts by 'Result Code of Interview', i.e., variable '*m11_q07*', from the Viet Nam data file '[STEP - Vietnam - Data for Weights - Revised Jan 2 2012.dta](#)'. The table reveals that, based on the 'result code of interview' assigned by the interviewers, there are 3405 acceptable interviews (i.e., codes 51 to 78) in the Viet Nam final data file.

Table 2: Result code of interview- Viet Nam Data File '[STEP - Vietnam - Data for Weights - Revised Jan 2 2013-1](#)'

Result Code & Description	Frequency	Percent
51-Refusal by SP to continue modules 2-7	1	.0
52-Interview termination-SP unusual circumstance	2	.0
61-SP unable to read language of GB	27	.4
62-SP refusal to begin GB	34	.5
63-SP unable to do GB due to unusual circumstance	1	.0
64-SP blindness or visual impairment-unable to do GB	1	.0
71-SP looked through GB-unable to read & write language of GB	12	.2
72-SP refused to continue GB after starting	35	.5
73-GB interruption too long	12	.2
74-GB terminated due to unusual circumstance	2	.0
76-Attempted part or all of General Booklet; less than 3 correct Core answers	132	1.9
81-No attempt at any of the exercise booklet items / refusal to start	50	.7
82-Interruption too long during exercise booklet administration	2	.0
83-Unusual circumstance, Exercise Booklet administration terminated	3	.0
84-Attempt at some Exercise Booklet items but no answers marked	24	.4
85-Attempt at all Exercise Booklet items but no answers marked	4	.1
86-Marked some Exercise Booklet items, not all items attempted	355	5.2
87-Marked some Exercise Booklet items, all items attempted	1034	15.2
88-Marked all Exercise Booklet items	1674	24.6
Sub-Total	3405	50.0
System Missing	3405	50.0
Total	6810	100.0

4.2 Revisions to Input Data File

The following revisions were made to the Viet Nam data file:

- 1) Renamed variables as follows,
 - a) RENAME VARIABLES (final_result_code=final_status)
 - b) RENAME VARIABLES (househol=HhldsPSU)
 - c) RENAME VARIABLES (replaced=HH_Replaced_Ind)
 - d) RENAME VARIABLES (urbrur=UrbRur_Ind)
- 2) Revised the value labels for the variables 'm11_q07', the Result of Interview Code, and 'final_status', the final status code for each case, to correspond to the Codes and code descriptions in Table 3 of this document.
- 3) 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.

- 4) **SPSS Revised File for Input to Weighting Process: 'W01-VNM Final Revised Data File(06-Mar-2013)'**
 - a) This file is the primary file for input to the process of calculating the survey weights.
 - i) Table 3 provides the list of key variables for weighting that are included in 'W01-VNM Final Revised Data File(31-Jan-2013)'

Table 3: Key Variables in Primary Input File for Weighting Process		
Variable Name	Source	Description
country	Sample File	Country abbreviation code VNM
hhn	Viet Nam Data File	Household Number-assigned in field
psusmpid	Sample File	PSU ID – sampled EAs
HhldsPSU	Sample File	Sample Frame # of households in PSU=EA
numhh	Viet Nam File 'sampleMarch_5'	Households-# in PSU, revised March 5, 2013
HHsmpSeq	Sample File	Household Sample Sequence #s 01 to 30
HHsmpType	Sample File	HH Sample Indicator
HHsmpOrder	Sample File	Sample activation order
m1c_a	Viet Nam Data File	NUMBER A-# of eligible persons
M11_Q7	Viet Nam Data File	Result code of interview
final_status	Viet Nam Data File	Final Status Code for sampled case
FinStat1	Created during file preparation	Revised Final Status Code for sampled case
v18_a	Viet Nam Data File;	Probability of Selection of EA; Frame variable 'District Probability'

- b) The final revised file for input to the weighting process includes all 6810 sampled cases. Table 4 provides the breakdown of the 6810 cases by Final Status Code 'FinStat1', and by Household Sample Type.
- i) For weighting purposes, a finalized case is a sample household for which a final status code has been assigned. During the file preparations for weighting, a final status code was assigned to each of the 6810 sampled cases.

Table 4: Viet Nam - Number of Finalized Cases by Final Status Code by HH Sample Type			
Final Status Code	Initial Sample	Reserve Sample	Total
01-Refusal-Household	93	65	158
02-Unusual circumstance-Household	9	4	13
03-No household contact after 3 visits	42	6	48
04-Temporarily absent/unavailable for field period	199	169	368
05-No competent household member to interview	58	28	86
06-Language problem	10	6	16
07- Dwelling could not be found / no household at selected address	21	6	27
08-Vacant Dwelling	83	58	141
09- Not habitable/destroyed/converted to business	46	30	76
51-Refusal by SP to continue modules 2-7	0	1	1
52- Unusual circumstance - SP (modules begun)	2	0	2
61-SP unable to read language of GB	24	3	27
62-SP refusal to begin GB	29	5	34
63-SP unable to do GB – unusual circumstance	1	0	1
64-SP unable to do GB – blindness or visual impairment	0	1	1
71-SP unable to read/write language of GB	10	2	12
72-SP refused to continue GB after starting	25	10	35
73-GB interruption	10	2	12
74-GB terminated due to unusual circumstance	2	0	2
76-SP marked some GB items; not all attempted	102	30	132
81-No attempt at any of the exercise booklet items / refusal to start	35	15	50
82-Interruption too long during exercise booklet administration	2	0	2
83-Unusual circumstance, Exercise Booklet administration terminated	3	0	3
84-Attempt at some Exercise Booklet items but no answers marked	19	5	24
85-Attempt at all Exercise Booklet items but no answers marked	2	2	4
86-Marked some Exercise Booklet items, not all items attempted	298	57	355
87-Marked some Exercise Booklet items, all items attempted	886	148	1034
88-Marked all Exercise Booklet items	1373	301	1674
94-Case dropped due to data quality issue	7	8	15
95-Sample case not activated	0	1054	1054
97- Household not interviewed-procedure error	14	1389	1403
Total	3405	3405	6810

4.3 Final Output Files from Weighting Process

The output file from the weighting process is,

- 1) [VNM-W01-STEP\(HH\) Weighted Data-Basic Weights \(06-Mar-2013\)-PRELIM](#)
 - a) This file contains the original data variables from the Viet Nam data file '[STEP - Vietnam - Data for Weights - Revised Jan 2 2012](#)' and the applicable PSU weights, household weights, person weights and other associated variables that were created during the weighting process.
 - b) The file includes 3405 participating households, each with a final household weight. The breakdown of the count of participating households is shown below in Table 6.
 - c) In addition to the original variables in the Viet Nam data file, the weighted data file includes other 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 5 provides the list of additional variables from the weighting process that are included in the final weighted data file.
 - ii) Table 6 provides the frequency count of result codes for participating households in the final weighted data file '[VNM-W01-STEP\(HH\) Weighted Data-Basic Weights \(06-Mar-2013\)-PRELIM](#)'.

Table 5: Variables Created or Used During Weighting Process - Added to Viet Nam Data File

Variable Name	Description	Values
FinStat1	Revised Final Status Code	Refer to Table 1 for code descriptions
FinRespCat_HH	Final Response Category-Households	1-Participant HH 2-Ineligible HH 3-Non-participant Eligible HH 4-Non-participant, activated HH; Unknown Eligibility 5-Non-participant, non-activated HH; Unknown Eligibility
FinRespCat_SP	Final Response Category-Selected Person	1-Participant SP 2-Non-participant; Eligible 3-Non-participant; Ineligible 7-Non-participant SP; non-participating activated HH 8-Non-participant SP; non-activated HH 9-Non-participant SP; Ineligible HH
W1_HHwt	Basic Phase 1 Household Weight	n/a
W_FinHHwt	Final Household Weight Use for estimation of Household characteristics	n/a
W_FinSPwt	Final Person Weight Use for estimation of Person characteristics	n/a

Table 6: Viet Nam Final Weighted Data File – Frequency of Final Status Codes FOR CASES WITH FINRESPCAT_HH=1, i.e., participating households		
	Frequency	Percent
51-Refusal by SP to continue modules 2-7	1	.0
52-Interview termination-SP unusual circumstance	2	.1
61-SP unable to read language of GB	27	.8
62-SP refusal to begin GB	34	1.0
63-SP unable to do GB due to unusual circumstance	1	.0
64-SP blindness or visual impairment-unable to do GB	1	.0
71-SP looked through GB-unable to read/write language of GB	12	.4
72-SP refused to continue GB after starting	35	1.0
73-GB interruption too long	12	.4
74-GB terminated - unusual circumstance	2	.1
76-Attempted part or all of General Booklet; <3 correct Core answers	132	3.9
81-No attempt at any exercise booklet items / refusal to start	50	1.5
82-Interruption too long during exercise booklet administration	2	.1
83-Unusual circumstance, Exercise Booklet administration terminated	3	.1
84-Attempt at some Exercise Booklet items, no answers marked	24	.7
85-Attempt at all Exercise Booklet items, no answers marked	4	.1
86-Marked some Exercise Booklet items, not all items attempted	355	10.4
87-Marked some Exercise Booklet items, all items attempted	1034	30.4
88-Marked all Exercise Booklet items	1674	49.2
Total	3405	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 '[W01-VNM Final Revised Data File\(06-Mar-2013\)](#)'.

5.1 Phase 1: PSU Weight

The Phase 1 probability of selection of each of the 1446 PSUs (i.e., EAs) that make up the '15% sample of EAs' was included in the sample frame of '15% sample of EAs' provided by the Viet Nam survey firm; the sample frame variable '*District Probability (fi)*' provided for each EA is the Phase 1 probability of selection of the EA in the '15% sample of EAs'.

For the first phase, the basic PSU weight is the inverse of the probability of selection of a PSU, which may be represented as,

$$1) \quad W_{1j} = \frac{1}{p_{1j}}$$

where,

p_{1j} - the probability of selection of the j^{th} EA to the '15% sample of EAs', i.e., the sample frame variable '*District Probability (fi)*'

5.2 Phase 1: Household Weight

Estimated Number of Households in Target Population

According to the Viet Nam survey firm, the sample frame variable '*District Probability (fi)*' provided for each EA is the Phase 1 probability of selection of the EA in the '15% sample of EAs'. Under this assumption, the estimated number of households in the target population is

$$\widehat{M}_1 = \sum_{j=1}^{1446} \frac{1}{p_{1j}} \times m_{1j} = 2,167,250$$

where,

$j = 1446$ is the number of EAs in the 15% sample of EAs,

p_{1j} - the probability of inclusion of the j^{th} EA in the '15% sample of EAs', i.e., the sample frame variable '*District Probability (fi)*'

m_{1j} - the sample frame number of households in each EA in the '15% sample of EAs',

$\widehat{M}_1 = 2,167,250$ - the estimated total number of households in the target population, i.e., in the urban EAs of Hanoi and Ho Chi Minh City.

Phase 1: Household Probability of Selection

The estimated probability of inclusion of a household in the '15% sample of EAs' is

$$p_{1jk} = \frac{\text{\# of Phase 1 Hhlds Selected}}{\text{\# of Hhlds in Target Population}} = \frac{\sum(\text{PSU Weight}) * (\text{\# of Hhlds in PSU})}{\text{\# of Hhlds in Target Population}}$$

$$= \frac{\sum_{j=1}^{1446} (W_{1j} * m_{1j})}{\widehat{M}_1}$$

The sample frame of the ‘15% sample of EAs’ includes a total of 189,701 households in 1446 EAs in the urban areas of Hanoi and Ho Chi Minh City.

The basic Phase 1 household weight for the households in the j^{th} PSU, W_{1jk} , is calculated as the inverse of its probability of selection,

$$2) \quad W_{1jk} = \frac{\widehat{M}_1}{\sum_j W_{1j} * m_{1j}} = \frac{2,167,250}{189,701} \approx 11.42$$

5.3 Phase 2: PSU Weight PSU Weight

The basic Phase 2 first-stage weight, i.e., the PSU weight, is based on a sample size of 227 PSUs. Although 250 PSUs were initially selected, only the 227 ‘initial’ PSU sample was actually activated during data collection; none of the 23 ‘reserve’ PSUs was used for the Viet Nam STEP data collection. Therefore, the STEP survey weight is based on the probability of selection for each of the 227 ‘initial’ PSU sample.

For the Viet Nam Phase 2 PPS sample, the probability of selection of the j^{th} PSU, say p_{2j} , may be represented as,

$$p_{2j} = \frac{y_2 * m_{2j}}{M_2}$$

where,

y_2 -the number of sampled PSUs from the 15% sample frame (i.e., $y_2=227$ CEAs in the case of the Viet Nam STEP Survey),

m_{2j} - the measure of size for the j^{th} PSU; in the case of the Viet Nam STEP Survey the measure of size for the first stage selection of PSUs was the sample frame number of households (i.e., weighted data file variable ‘**HhldsPSU**’) in each PSU in the Phase 2 sample frame,

M_2 -the total number of households in the Phase 2 sample frame is represented by

$$M_2 = \sum_{j=1}^Y m_{1j}$$

where,

Y - the total number of PSUs in the Phase 2 sample frame; there are 1446 PSUs in the Phase 2 sample frame.

$M_2 = 189,701$ is the Phase 2 sample frame total count of households for the 1446 EAs in the Phase 2 sample frame.

The basic Phase 2 first stage weight for the j^{th} PSU, W_{2j} , is calculated as the inverse of its probability of selection:

$$3) \quad W_{2j} = \frac{1}{p_{2j}} = \frac{M_2}{y_2 * m_{2j}}$$

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

Example 1: First Stage Weight

The basic Phase 2 first stage weight for the Viet Nam PSU # 126 which contains 88 sample frame households is calculated as follows:

$$W_{2j} = \frac{1}{p_{2j}} = \frac{M_2}{y_2 * m_{2j}} = \frac{189,701}{227 * 88} \approx 9.5$$

The basic Phase 2 first stage weight is similarly calculated for each of the 227 PSUs.

5.4 Phase 2: Basic Household Weight for Each PSU

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

$$4) \quad W_{2jk} = \frac{m'_{2j}}{n_{2j}} = \frac{m'_{2j}}{30}$$

where,

m'_{2j} - the number of households in the j^{th} PSU (*revised 05-March-2013, i.e., variable 'numhh'*) at the time of household selection,

n_{2j} - the number of households sampled in the j^{th} PSU, (i.e., n_{2j} = 30 households in each Viet Nam sampled PSU).

Example 2: Phase 2 Second Stage Basic Weight

The basic Phase 2 second stage weight for each of the 30 sampled households in the Viet Nam PSU # 126 which contains 88 households is calculated as follows:

$$W_{2jk} = \frac{m'_{2j}}{n_{2j}} = \frac{71}{30} \approx 2.367$$

The basic second stage weight is similarly calculated for each of the 227 PSUs.

5.4.1 Phase 2: 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 - 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 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
Individual modules 2-7 not begun	
32	Refusal by Selected Person to begin modules 2-7
33	Unusual circumstance - Selected Person (e.g., death in family, illness, fire in dwelling, etc.)
34	Refusal - Another household member refused to let selected individual do individual modules
35	Unable to contact Selected Person after three visits during field period
36	Temporarily absence/unavailability of Selected Person for field period (info from household member)
37	Hearing Problem - Selected Person is deaf or so hard-of-hearing, no translator available
38	Learning or mental disability (includes emotional conditions like severe depression) of Selected Person
39	Speech impairment of Selected Person prevented participation, no translator available
40	Language problem – Selected Person's language not understood by survey team, no translator available
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
10	Household moved away
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
94	Case dropped due to data quality issue

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
97	Household not interviewed due to field procedure error – wrong household selected in its place

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

$$5) \quad n_{2j} = n_{2jk} + n_{2jx} + n_{2jz} + n_{2ju}$$

where,

- n_{2j} - the number of households sampled in the j^{th} PSU, (i.e., $n_j = 30$ households in each Viet Nam sampled PSU),
- n_{2jk} - the number of households ‘k’ that participate in the j^{th} PSU,
- n_{2jx} - the number of ineligible households ‘x’ in the j^{th} PSU,
- n_{2jz} - the number of non-participating households ‘z’ with known eligibility in the j^{th} PSU,
- n_{2ju} - the number of non-participating households ‘u’ with unknown eligibility in the j^{th} PSU.

Table 7 provides the STEP counts from the Viet Nam data file by the outcome categories described above.

Table 7: Sample Count of Households by Final Response Category		
Final Response Category	Frequency	Percent
1-Participant HH (n_{2jk})	3405	50.0%
2-Ineligible HH (n_{2jx})	244	3.6%
3-Non-participant, Eligible HH (n_{2jz})	15	0.2%
4-Non-participant, Unknown Eligibility HH (n_{2ju})	3146	46.2%
Total (n_{2j})	6810	100.0

5.4.2 Phase 2: Household Weight - Adjustment for Unknown Eligibility

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

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

$$6) \quad \varepsilon = \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} = \frac{n_{2jk} + n_{2jz}}{n_{2jk} + n_{2jx} + n_{2jz}}$$

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

Example 3: Eligible Proportion

The counts in Table 7 provide input for calculation of the proportion of households with known eligibility.

$$\begin{aligned}
 \varepsilon &= \frac{\text{Total \# of known eligible cases}}{\text{Total \# of known eligible \& known ineligible cases}} \\
 &= \frac{n_{2jk} + n_{2jz}}{n_{2jk} + n_{2jx} + n_{2jz}} \\
 &= \frac{3405 + 15}{3405 + 244 + 15} = \frac{3420}{3664} \approx 0.933406
 \end{aligned}$$

5.4.3 Phase 2: Adjustment of Household Weights for Non-participation

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

$$7) \quad R_{2j} = \frac{\text{Total eligible households in PSU 'j'}}{\text{Total participating households in PSU 'j'}} = \frac{n_{2jk} + n_{2jz} + (\varepsilon * n_{2ju})}{n_{2jk}}$$

➤ The response adjustment ratio, R_{2j} , incorporates the adjustment for unknown eligibility, ε .

- 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 #126

n_{2j}	n_{2jk}	n_{2jx}	n_{2jz}	n_{2ju}
30	15	0	0	15

Therefore,

$$\begin{aligned}
 R_{126} &= \frac{\text{Total eligible households in PSU } j}{\text{Total participating households in PSU } j} = \frac{n_{2jk} + n_{2jz} + (\epsilon * n_{2ju})}{n_{2jk}} \\
 &= \frac{15 + 0 + (.933406 * 15)}{15} \\
 &\approx 1.933406
 \end{aligned}$$

The response adjustment ratio, R_{2j} , is similarly calculated for all 227 sampled PSUs.

5.4.4 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 Phase 1 household weight, W_{1jk} , and the Phase 2 PSU weight, W_{2j} , the Phase 2 basic household weight, W_{2jk} , and the Phase 2 response adjustment, R_{2j} , for the basic household weight. Therefore, for household 'k' in PSU 'j', the final household weight, W_H (i.e., weighted data file variable '**W_FinHHwt**'), is calculated as follows:

$$8) \quad W_H = W_{1j} * W_{2j} * W_{2jk} * R_{2j}$$

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

Example 5: Second Stage – Final Household Weight

For PSU #126,

$$\begin{aligned}
 W_H &= W_{1j} * W_{2j} * W_{2jk} * R_{2j} \\
 &= 11.42 \times 9.496 \times 2.367 \times 1.933406 \\
 &\approx 496.433
 \end{aligned}$$

The household weight is similarly calculated for all sampled households.

5.4.5 Weighted Household Counts

Based on the information in the sample frame '[VNM Sample Frame-HN-HCN-20-April-2012](#)' it is estimated that there are 2,167,250 households in the Viet Nam target population comprised of urban areas of Hanoi and Ho Chi Minh City.

Table 8 provides,

- i. the weighted count of households based on the sample frame count of households for the 227 sampled PSUs;
- ii. the weighted count of households based on the final status of cases in the Viet Nam data file data.

Table 8: Viet Nam – Weighted Number of Households – Sample Frame vs. STEP Data File		
Source of PSU Household Count	# of Sampled PSUs	Weighted Sum of Households
Estimate based on # of Households in Sample Frame ' VNM Sample Frame-HN-HCN-20-April-2012 '	1446	2,167,250
Based on # of Households in STEP Data File ' STEP - Vietnam - Data for Weights - Revised Jan 2 2012 ' [all participants, i.e., participants, non-participants, ineligible]	227	2,167,250
Based on # of Households in STEP Data File ' STEP - Vietnam - Data for Weights - Revised Jan 2 2012 ' [eligible households only]	227	1,969,989

As shown in the third row of Table 8, when the 244 ineligible households are excluded, the weighted number of eligible households is 1,969,989.

5.5 Phase 2: Person Weight

The basic third stage Phase 2 weight for the i^{th} person, $W3_{2jki}$, is calculated as follows:

$$9) \quad W3_{2jki} = \frac{a_{2jk}}{b_{2jk}} = \frac{a_{2jk}}{1} = a_{2k}$$

where,

a_{2jk} - the number of eligible persons in the k^{th} household in the j^{th} PSU,

b_{2jk} - the number of persons sampled in the k^{th} household in the j^{th} PSU, (i.e., $n_{jk} = 1$ person in each Viet Nam sampled household).

$$\text{N.B.} \quad A_{2jk} = \sum_{k=1}^{n_{2jk}} a_{2jk} = \text{total \# of selected persons in the } j^{\text{th}} \text{ PSU}$$

where,

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

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

For PSU #126, selected household # 03 has 5 eligible persons. Therefore,

$$W3_{2jki} = \frac{a_{2jk}}{b_{2jk}} = \frac{5}{1} = 5$$

5.5.1 Phase 2: 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, 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

The third stage person weight is only relevant for sampled households in the final response category $\text{FinRespCat}=1$, i.e., participating households. Table 9 provides the pertinent frequency counts of the Final Result Codes.

Table 9: Final Status Code for Participating Households (i.e., Households with $\text{FinRespCat_HH}=1$)

Final Status Code & Description	Frequency	Percent
51-Refusal by SP to continue modules 2-7	1	.0
52-Interview termination-SP unusual circumstance	2	.1
61-SP unable to read language of GB	27	.8
62-SP refusal to begin GB	34	1.0
63-SP unable to do GB due to unusual circumstance	1	.0
64-SP blindness or visual impairment-unable to do GB	1	.0
71-SP looked through GB-unable to read/write language of GB	12	.4
72-SP refused to continue GB after starting	35	1.0
73-GB interruption too long	12	.4
74-GB terminated - unusual circumstance	2	.1
76-Attempted part or all of General Booklet; <3 correct Core answers	132	3.9
81-No attempt at any exercise booklet items / refusal to start	50	1.5
82-Interruption too long during exercise booklet administration	2	.1
83-Unusual circumstance, Exercise Booklet administration terminated	3	.1
84-Attempt at some Exercise Booklet items, no answers marked	24	.7
85-Attempt at all Exercise Booklet items, no answers marked	4	.1
86-Marked some Exercise Booklet items, not all items attempted	355	10.4
87-Marked some Exercise Booklet items, all items attempted	1034	30.4
88-Marked all Exercise Booklet items	1674	49.2
Total	3405	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, $W3_{2jki}$, 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:

$$10) \quad A_{2jk} = A_{2jki} + A_{2jkq}$$

where,

$$A_{2jk} = \sum_{k=1}^{n_{2jk}} a_{2jk} \quad - \quad \text{the total number of eligible persons for all } n_{jk} \text{ participating households in the } j^{\text{th}} \text{ PSU,}$$

$A_{2jki} = \sum_{k=1}^{n_{2jk}} a_{2jki}$ - the total number of eligible persons 'i' that participate for all n_{jk} participating households in the j^{th} PSU,

$A_{2jkq} = \sum_{k=1}^{n_{2jk}} a_{2jkq}$ - the total number of non-participating eligible persons 'q' for all n_{jk} participating households in the j^{th} PSU,

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

$$11) \quad R_{2jk} = \frac{\text{Total eligible persons in all participating sampled households in PSU 'j'}}{\text{Total participating persons in all participating sampled households in PSU 'j'}} = \frac{A_{2jk}}{A_{2jki}}$$

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 #126, there are 38 eligible persons across all participating households. Also, all selected persons participated in the survey and thus the selected persons represent the 38 eligible persons in the participating households. Therefore,

$$A_{2jk} = \sum_{k=1}^{n_{2jk}} a_{2jk} = 38; \quad A_{2jki} = \sum_{k=1}^{n_{2jk}} a_{2jki} = 38$$

$$R_{2jk} = \frac{A_{2jk}}{A_{2jki}} = 38 = 1.00$$

Table 10 provides the sample counts of the Final Response Category for selected persons. This table shows the overall response rate, i.e., 61.8%, for the Viet nam Step. Note that all selected eligible persons in participating households were STEP participants. In other words, although there were non-responding eligible households, there were no non-responding selected persons within the participating households.

Table 10: Sample Counts by Selected Person Final Response Category for Participating Households		
For Participating Households		
Final Response Category-SP	Frequency	Percent
1-Participant SP	3405	61.8
7-Non-participant SP; non-participating activated HH	2107	38.2
Total	5512	100.0

Based on the above table, there are 3405 participating selected persons for which a final non-zero person weight was calculated.

5.5.2 Phase 2: Final 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 Phase 1 PSU weight, Phase 1 household weight, and the Phase 2 PSU weight, and the response adjustment for the basic household weight), the basic third stage person weight, and the response adjustment for the basic person weight. Therefore, for the selected person 'i' in household 'k' in PSU 'j', the final person weight, W_{SP} (i.e., weighted data file variable 'W_FinSPwt'), is calculated as follows:

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

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

Example 8: Third Stage – Final Person Weight

For PSU #126, selected household # 03 with 5 eligible persons, the selected person was a participant. All selected persons participated in PSU #126, therefore, $R_{jk} = 1$

$$W_{SP} = 496.433 * 3 * 1 \approx 1489.299$$

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

5.5.3 Weighted Person Counts

The weighted count of the non-institutionalized persons 15 to 64 years of age (inclusive) living in private dwellings in urban areas of **Hanoi and Ho Chi Minh City** at the time of data collection is provided in Table 11. Based on the 3405 participating selected persons, the sum of the final person weights from the STEP survey is 6,120,368.

Table 11: Sample Frequency & Weighted Persons Totals by Final Selected Person Response Category		
For Participating Households		
Final Response Category-Selected Person	# of Sampled Cases	Weighted Total # of Eligible Persons in Target Population
1-Participant SP	3,405	6,120,368
7-Non-participant SP; non-participating activated HH	2107	0
8- Non-participant SP; non-activated HH	1054	0
9- Non-participant SP; Ineligible HH	244	0
Total	6,810	6,120,368

For participating selected persons, Table 12 provides a summary of the central tendency of the number of eligible persons in the participating households. Based on the average of 3.1 eligible persons per participating household and an estimated 1,969,989 households in the target population, there are approximately 6,100,000 persons aged 15 to 64 in the target population.

Table 12: Summary of Average # of Eligible Persons, Average Person Weights & Average Household Weights			
Statistic	NUMBER A- # of eligible persons	Final Person Weight	Final Household Weight
	3405	3405	3405
Mean	3.1137	1,797.46	578.56
Std. Error of Mean	.02469	17.35	2.97
Median	3.0	1,623.50	560.60
Mode	2.0	1,230.59	615.30
Minimum	1	265.48	149.43
Maximum	12	10,961.12	1,826.85
Sum	10,602	6,120,367.87	1,969,989.15