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

UKRAINE

July 5, 2014

Table of Contents

1   Survey Design Overview

The Ukraine 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 obtain 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 Ukraine 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
The sample excluded individuals permanently institutionalized in medical facilities, military quarters, and prisons; these exclusions totalled about 725,000 persons or about 2% of the population. Also excluded from STEP was a 30-km zone around the Chernobyl Nuclear Power Plant, an area with a high level of radiation contamination where public access is restricted and all population was evacuated.

1.2 Sample Frame

There are two sample frame components from which the STEP sample was selected:

1) ULMS Panel
2) Voting Precincts: The list of 300 urban Voting Precincts covering all strata from the Ukraine Global Adult Tobacco Survey (GATS), 2010.

The sample frames of the ULMS Panel and the STEP Voting Precincts are independent of each other.

The most recent Ukrainian Census was conducted in 2001. Table 1 provides the 2001 Census population by stratum and the 2011 demographic estimate of the target population (i.e., population aged 15 to 64) by stratum.

Table 1: Urban Population (2001) & Estimated Target Population (2011) by Stratum		
Stratum **Urban part of administrative unit**	**Urban Population (2001)** [*] **All ages**	**Estimated Population (2011)** [**] **Aged 15-64**
Crimea	1,632,365	1,129,651
Kyiv (city)	2,611,327	2,060,273
Kyivska	1,053,486	767,588
Vinnytska	818,910	588,775
Volynska	533,171	384,445
Dnipropetrovska	2,960,251	1,989,900
Donetska	4,363,555	2,846,737
Zhytomyrska	775,377	531,907
Zakarpatska	466,004	328,509
Zaporizka	1,458,216	999,478
Ivano-Frankivska	592,995	430,654
Kirovogradska	681,980	443,460
Luhanska	2,190,744	1,432,774
Lvivska	1,558,747	1,105,006
Mykolayivska	838,828	580,115
Odeska	1,624,609	1,144,495
Poltavska	956,756	651,620
Rivnenska	549,727	403,571
Sumska	842,897	574,792
Ternopilska	485,560	346,697
Kharkivska	2,288,681	1,617,580
Khersonska	706,251	474,975
Khmelnitska	729,575	534,806
Cherkaska	753,627	523,252
Chernivetska	373,520	278,326
Chernihivska	727,212	499,920
Total	**32,574,371**	**22,669,306**

* Source: Census
** Source: State Statistics Service of Ukraine

1.3 Sample Design

The Step sample design is outlined in the document "*STEP Skills Measurement Study + UKRAINIAN LONGITUDINAL MONITORING SURVEY 2012: Technical Report*", which was prepared by the Ukrainian survey team.

The Ukraine used a stratified sample design comprised of three components:
1) ULMS Panel Sample;
2) New ULMS-2012 subsample;
3) Step Urban Subsample.

In the ULMS Panel Sample and the Step Urban Subsample, the urban sample was selected within 26 strata consisting of the Autonomous Republic of Crimea, the city of Kiev, and 24 Oblasts, i.e., geographic administrative units. In the New ULMS-2012 subsample, there was no explicit stratification.

1.3.1 ULMS Panel Sample Design

In the Ukraine STEP survey, the ULMS Panel component "*subsample consisted of the respondents included in the sample of baseline research ULMS-2003 and follow up surveys ULMS-2004 and 2007* [1]"; essentially, it followed the sample design of the ULMS 2003 Panel Study, which was a stratified four-stage sample design.

"*Stratification of the population was carried out in order to adequately reflect the main features of administrative-territorial division of Ukraine in the sample, as well as to provide the selection of the units from more homogenous groups by the main characteristics. 51 strata were singled out, which represented all the regions of Ukraine with division by urban and rural area (25 oblasts and Kyiv city).* [1]"
The STEP survey was limited to the urban area of Ukraine.

First Stage Sample Unit - ULMS 2003 Panel
The ULMS 2003 Panel primary sample unit and thus the STEP primary sample unit (PSU) was an urban settlement. The PSUs for the ULMS 2003 Panel were selected using a probability proportional to size (PPS) sampling method, where the measure of size was the population size (i.e., number of persons aged 15 to 72 [2]) in a PSU.

"*Selection of the first stage was implemented independently in every stratum. Within one oblast, for urban and rural areas corresponding lists of urban settlements and rural rajons were constructed sorted by descending population size. Then there was implemented the selection procedure with probability proportional to the size.* [3]"

Second Stage Sample Unit - ULMS 2003 Panel
The ULMS 2003 Panel secondary sample unit and thus the STEP second stage sample unit (SSU) was a Postal District. The SSUs for the ULMS 2003 Panel were randomly in each sampled PSU.

"*Within each PSU, a list of residential postal districts is constructed. Sometimes this can be done centrally; in other cases a fieldworker must consult local authorities to construct the necessary list in a given PSU. Once the list is constructed, a few postal districts are randomly chosen taking into account the*

---

[1] Page 6, *STEP Skills Measurement Study + UKRAINIAN LONGITUDINAL MONITORING SURVEY 2012: Technical Report*
[2] ULMS Panel target population of persons aged 15-72 differs from the STEP target population of persons aged 15 to 64.
[3] Page 20, *STEP Skills Measurement Study + UKRAINIAN LONGITUDINAL MONITORING SURVEY 2012: Technical Report*

number of residential routes, which is proportional to the number of apartments. The number of selected postal units depends on the size of the ultimate clusters that we want.[4]"

Third Stage Sample - ULMS 2003 Panel
The ULMS 2003 Panel third stage sample unit and thus the third stage sample unit for STEP was a household. "*For selecting the households by every selected postal district the street, building, and households' addresses were randomly selected one after another.* [3]"

Fourth Stage Sample - ULMS 2003 Panel
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.

1.3.2   ULMS 2012 Subsample Design

For the ULMS 2012 Subsample component, the Ukraine STEP survey employed a sample design that was similar to the sample design of the ULMS 2003 Panel Study, which was a stratified four-stage sample design; the stratification was the same as described above for the ULMS 2003 Panel component. Although the ULMS 2012 Subsample component used the same four stage sample design as the ULMS 2003 Panel Study, there was a variation in the selection of the second stage sample units whereby a subsample of the sampled SSUs from the ULMS 2003 Panel Study was systematically selected.

First Stage Sample Unit - ULMS 2012 Subsample
The first stage sample for the ULMS 2012 Subsample component used the same Settlement PSUs as the ULMS 2003 Panel.

Second Stage Sample Unit - ULMS 2012 Subsample
The ULMS 2012 Subsample component's secondary sample unit was a Postal District; for the ULMS 2012 Subsample "*170 sampling points have been systematically selected from the complete list of 406 panel SSUs* [5]".

It should be noted that the STEP target population comprised only urban areas; 280 SSUs out of "*the complete list of 406 panel SSUs*" were urban SSUs, and 115 urban SSUs were included in the "*170 sampling points*".

Third Stage Sample - ULMS 2012 Subsample
The ULMS 2003 Panel third stage sample unit and thus the third stage sample unit for STEP was a household; "*within given SSU, interviewer received the last panel address and prolonged the route selecting each 5th household for interview. The interviewer had to interview 3 households having eligible individual respondents.* [3]"

Fourth Stage Sample - ULMS 2012 Subsample
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.

---

[4] Page 18, *STEP Skills Measurement Study + UKRAINIAN LONGITUDINAL MONITORING SURVEY 2012: Technical Report*
[5] Page 6, *STEP Skills Measurement Study + UKRAINIAN LONGITUDINAL MONITORING SURVEY 2012: Technical Report*

1.3.3   STEP Urban Subsample Design

The Step Urban Subsample component used a 3 stage sample design.

First Stage Sample Unit - STEP Urban Subsample
The STEP Urban Subsample component's primary sample unit (PSU) was a Voting Precinct; 70 Voting Precincts were systematically selected from the 300 Voting Precincts that had been selected for the Ukraine Global Adult Tobacco Survey (GATS), 2010.

The following summarizes the GATS sample design:

"*The 2010 Ukraine GATS used a stratified two-stage sample design. At the first stage, primary sampling units (PSUs) were selected randomly by probability proportional to size. Definition of PSU differed for urban and rural areas. For the urban population, PSUs were equal to voting precincts defined by Central Electoral Commission for 2007 nationwide parliamentary elections, and those PSUs consisted of eligible voters (citizens of 18 years old and older). For the rural population, PSUs were equal to villages (or groups of small villages), and those PSUs consisted of the total population from the 2001 Population Census. There were 300 PSUs selected in the urban areas and 300 PSUs selected in the rural areas.* [6]"

Second Stage Sample Unit - STEP Urban Subsample
 The STEP Urban Subsample component's secondary 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.

Third Stage Sample - STEP Urban Subsample
 The STEP Urban Subsample 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.

---

[6] Survey Report for *Global Adult Tobacco Survey (GATS), Ukraine 2010*

1.4 Sample Size

The Ukraine firm's sampling objective was to obtain interviews from a minimum of 2400 individuals in the urban areas of the country. Table 2 provides the number of PSUs and the urban household sample size by Subsample component.

Table 2: Ukraine 'Urban' Sample: Number of PSUs[7] & Number of HHs By Oblast

Oblast	ULMS-2003 panel subsample		New ULMS-2012 subsample		STEP urban subsample		Total	
	PSUs	HHs	PSUs	HHs	PSUs	HHs	PSUs	HHs
1	15	105	6	20	3	90	24	215
2	20	147	9	57	6	180	35	384
3	8	55	3	13	2	60	13	128
4	8	64	3	10	2	60	13	134
5	4	17	2	9	1	30	7	56
6	25	185	10	45	6	180	41	410
7	43	392	17	102	10	300	70	794
8	4	33	2	7	2	60	8	100
9	1	7	1	15	1	30	3	52
10	13	122	5	35	3	90	21	247
11	9	61	4	22	1	30	14	113
12	6	63	2	6	1	30	9	99
13	22	212	9	42	5	150	36	404
14	13	92	5	39	3	90	21	221
15	7	67	3	13	2	60	12	140
16	12	78	5	30	4	120	21	228
17	12	82	5	19	2	60	19	161
18	4	39	1	4	1	30	6	73
19	8	70	3	13	2	60	13	143
20	2	9	2	13	1	30	5	52
21	19	204	8	51	5	150	32	405
22	8	59	3	26	1	30	12	115
23	6	39	2	11	2	60	10	110
24	6	64	2	11	1	30	9	105
25	3	23			1	30	4	53
26	9	78	4	20	2	60	15	158
Total	279	2367	115	633	70	2100	473	5100

---

[7] Source: Excel File 'STEP_panel_part_with 2003 prob_1-Revd 24-Jun-2014-1'

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 Ukraine final data file; a '**final_status**' was assigned to each sampled case.

Table 3:	*Full Literacy Assessment* 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
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
11	Panel HH has moved; not located
12	Panel HH has united with other panel HH
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, originally sampled in Rural area
95	Sample case justifiably not activated, e.g., reserve sample household not needed , sample goal achieved
96	General non-response; Reason unspecified by survey firm
97	Household not interviewed due to sample selection error – wrong household selected in its place
II. Person-level Codes	
Code	**Code Description**
	Individual modules 2-7 not begun or unacceptable interview
32	Refusal by Selected Person to begin modules 2-7
33	Unusual circumstance - Selected Person (e.g., death in family, illness, fire in dwelling, etc.)
34	Refusal - Another household member refused to let selected individual do individual modules
35	Unable to contact Selected Person after three visits during field period
36	Temporarily absence/unavailability of Selected Person for field period (info from household member)
37	Hearing Problem - Selected Person is deaf or so hard-of-hearing, no translator available
38	Learning or mental disability (includes emotional conditions like severe depression) of Selected Person
39	Speech impairment of Selected Person prevented participation, no translator available
40	Language problem – Selected Person's language not understood by survey team, no translator available
98	Sample selection error – an incorrect eligible person was selected
99	Sample selection error – an ineligible person was selected
	Individual modules 2-7 begun but not completed
51	Refusal by Selected Person to continue modules 2-7
52	Unusual circumstance - Selected Person
	General Booklet (module 9) not begun
61	Unable to read or write in language of General Booklet – Selected Person refused to begin
62	Refusal by Selected Person to begin General Booklet
63	Unusual circumstance - Selected Person
64	Blindness or visual impairment of Selected Person prevented completion of General Booklet

Table 3:	*Full Literacy Assessment* 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
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

The Ukraine sample design resulted in differential probabilities of selection for the selected persons. Consequently, each selected person in the survey does not necessarily represent the same number of persons in the target population. To account for differential probabilities of selection due to the nature of the design and to ensure accurate survey estimates, 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 Ukraine 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 each activated PSU;
   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 Ukraine 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 Ukraine - BQ Data for weighting_s7.dta' provided by the WB core team. The file 'STEP Ukraine - BQ Data for weighting_s7.dta' contains a data record for each of the 5100 sampled urban households from the Ukraine sample frames. An SPSS version of this file, i.e., 'STEP Ukraine - BQ 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 Ukraine SPSS data file 'STEP Ukraine - BQ Data for weighting_s7':

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

   a) '*final_status*' revised as follows:
      i)   IF $((m11\_q07 \neq 0)$ & (field_rst = 1 | field_rst = 3)) final_status=m11_q07.
      ii)   IF $((m11\_q07 = 0)$ & (field_rst = 1 | field_rst = 3)) final_status=31.
      iii) IF (field_rst = 1 | field_rst = 3) Final_status=Final_status.
      iv)   IF (Urban_Rural_Ind = 1 & TYPE_Chng = 1) Final_status=94.
      v)   IF (field_rst = 2) Final_status=95.
      vi)   IF (field_rst = 4) Final_status=97.
      vii) IF (field_rst = 5) Final_status=98.
      viii) IF (field_rst = 6) Final_status=99.

   b) 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.

2) '*Urban_Rural_Ind*', a Urban Rural Indicator; created as follows:

    a) RECODE 'Type_set' (1=2) (MISSING=9) (ELSE=1) INTO 'Urban_Rural_Ind'.
    b) VALUE LABELS for 'Urban_Rural_Ind'
        i) '1-URBAN'
        ii) '2-RURAL'

3) 'TYPE_Chng': Indicator that a case that was originally sampled in a rural (or urban) area moved to an urban (or rural) area; created as follows:
    a) IF (Urban_Rural_Ind $\neq$ TYPE_Orig) TYPE_Chng=1.
    b) IF (Urban_Rural_Ind = TYPE_Orig) TYPE_Chng=0.
    c) VALUE LABELS for 'TYPE_Chng'
        i) ' 0-HH in URBAN_RURAL Area as Original TYPE '
        ii) ' 1-HH Moved to Different URBAN_RURAL Area than Original TYPE '

4) The following variables were added to the final BQ data file:
    a) 'PSU_CODE': PSU ID, 2003
    b) 'PD_CODE': Postal District ID, 2003
    c) 'PDs_Total2012': Postal Districts 2012, Number of Postal Districts within a PSU in 2012
    d) 'PDs_Smpd2012': Sampled Postal Districts 2012, Number of Sampled Postal Districts within a PSU in 2012
    e) 'PDs_Total2003E': Postal Districts 2003, Estimated Number of Postal Districts within a PSU in 2003
    f) 'PDs_Smpd2003E': Sampled Postal Districts 2003, Estimated Number of Sampled Postal Districts within a PSU in 2003

5) 'HHs_Total': Total number of Households in a current SSU (or PSU for Subsample 3)
    a) For Subsamples 1 & 2, 'panel_list_last_with 2003 prob' which was provided by the Ukraine Survey Team.
    b) For Subsample 3, the Total number of Households was deduced from the Excel file 'enumerated_PSUs_clean' which was provided by the Ukraine Survey Team.

6) 'HHs_PDsplit': Revised Total number of Households: created as follows:
    a) For Subsamples 1 & 2, 'HHs_Total1' is the result of dividing the variable 'HHs_Total' by the number of PSU_IDs in a Postal District, i.e., in an SSU.
        i) Rationale: Subsequent to the sampling of SSUs for the ULMS 2003 Panel, some of the Postal Districts have been merged; it seems that the variable 'HHs_Total' reflects the total number of Households in the current (sometimes merged) Postal Districts. However, it appears that the number of PSU_IDs is the same as that of the ULMS 2003 Panel; therefore, it is hypothesized that the total number of Households in the current (sometimes merged) Postal Districts is the total number of households across all PSU_IDs in the Postal District.

7) Included interim weights provided by the Ukraine Team as follows:

    a) For Subsample=1, included Household weights from ULMS 2007, i.e., variable 'HHwt_2007'. These weights were provided by the Ukraine Team in the Excel file 'HH_2007'; an ULMS 2007 Household Weight was included for each variable 'HH_ID'.
    b) For Subsample=2, included estimated 2003 weights, i.e., variable 'PSUSSU_EstWt', based on the estimated probabilities of selection of a PSU and SSU. These estimated 2003 weights are based on the information provided by the Ukraine Survey Team in the Excel file 'STEP_panel_part_with

2003 prob_1-Revd 24-Jun-2014-1'; an ULMS 2003 PSU_SSU Weight Estimate was included for each variable 'PSU_ID'.

c) For Subsample=3, included the initial Voting Precinct weights, i.e., variable 'GATS_VP_Wgt', from Ukraine Global Adult Tobacco Survey (GATS), 2010. These GATS Voting Precinct weights were provided by the Ukraine Survey Team in the Excel file 'STEP urban subsample'; column H, "Basic weight, w1", provided the GATS Voting Precinct weight for the 300 Voting Precincts that were selected for GATS, while column C, "PSU_ID (for STEP)" indicated the Voting Precincts that were sub-sampled for STEP.

8) Excluded 54 cases of Subsample=1 from the weighting process; these cases were actually originally sampled from a RURAL settlement and thus they are not eligible for the STEP Survey of URBAN households.
    a) The 54 cases were identified as 'Urban_Rural_Ind = 1 & TYPE_Chng = 1'; the variable TYPE_Chng was created during the weighting process to identify cases that changed the type of area since the 2003 ULMS Panel.

4.1.1.1 Input File – Revised for Weighting

The SPSS Revised Input File for weighting is 'W00-1_R-UKR Final STEP URBAN Data File(24-Jun-2014)-for wtg'. This file is the primary file for the process of creating the survey weights; it includes 5044 urban sampled cases; each case includes a final status code. Table 4 provides the list of key variables for weighting purposes that are included in the file.

Table 4: Key Variables for Weighting Process in Primary Input File 'W00-1_R-UKR Final STEP URBAN Data File(24-Jun-2014)-for wtg'		
Variable Name	**Source**	**Description**
Country	Sample File	Country abbreviation code
Case_Num	Created during file preparation	Case Number: 24-Jan-2014 Data File
PSU_ID	Final Data File	PSU Sample ID#
HH_ID	Final Data File	Household ID
Subsample	Final Data File	Sample Component
OBLAST	Final Data File	Stratum
OBLAST_Orig	File 'STEP_panel_part_with 2003 prob_1-Revd 24-Jun-2014-1' provided by Ukraine	OBLAST ID #, ULMS 2003 Panel
Type_set	Final Data File	Type and Size of Settlement
Urban_Rural_Ind	Created during file preparation	Sampled case ID#
TYPE_Orig	File 'STEP_panel_part_with 2003 prob_1-Revd 24-Jun-2014-1' provided by Ukraine	TYPE of Area (2003), Urban-Rural
PSU_CODE	File 'STEP_panel_part_with 2003 prob_1-Revd 24-Jun-2014-1' provided by Ukraine	PSU ID, 2003
HHsmp_Order	Sample File	Sample activation order
HHsmp_Type	Sample File	HH Sample Indicator
age	Final Data File	age of selected individual
m1c_a_clean	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	Created during file preparation	Case Final Result Code
HHs_Total	• For Subsamples 1 & 2, Excel file 'panel_list_last_with 2003 prob' • For Subsample 3 Excel file 'enumerated_PSUs_clean'	# of listed households in PSU or SSU
HHs_PDsplit	File 'STEP_panel_part_with 2003 prob_1-Revd 24-Jun-2014-1' provided by Ukraine	Estimated # of Households in 2003 Selected Postal District
PDs_Total2012	File 'STEP_panel_part_with 2003 prob_1-Revd 24-Jun-2014-1' provided by Ukraine	Postal Districts 2012 Population Size within PSU
PDs_Smpd2012	File 'STEP_panel_part_with 2003 prob_1-Revd 24-Jun-2014-1' provided by Ukraine	Postal Districts 2012 Sample Size within PSU
PDs_Total2003E	File 'STEP_panel_part_with 2003 prob_1-Revd 24-Jun-2014-1' provided by Ukraine	Postal Districts 2003 Population Size Estimate within PSU
PDs_Smpd2003E	File 'STEP_panel_part_with 2003 prob_1-Revd 24-Jun-2014-1' provided by Ukraine	Postal Districts 2003 Sample Size Estimate within PSU
HHwt_2007	File 'HH_2007' provided by Ukraine	ULMS 2007 Final Household Weight
PSUSSU_EstWt	File 'STEP_panel_part_with 2003 prob_1-Revd 24-Jun-2014-1' provided by Ukraine	PSU_SSU Estimate of Weight, 2003
GATS_VP_Wgt	File 'STEP urban subsample' provided by Ukraine	GATS Voting Precinct weight

Table 5 provides for each subsample component the breakdown of the 5046 urban sample cases by the final result of the interview attempt, i.e., by **'*final_status*'** code. Note that there were originally 5100 urban sampled cases, but 54 urban cases of Subsample=1 were excluded from the weighting process; these cases were actually originally sampled from a Rural settlement and thus they are not eligible for the STEP Survey of Urban households.

Table 5: Ukraine - Number of Sampled Households by 'final_status'				
	Number of Cases			
Final Status	**1-ULMS 2003 Panel**	**2-New ULMS 2012**	**3-STEP Urban Subsample**	**Total**
01-Refusal-Household	177	143	604	**924**
02-Unusual circumstance-Household	11		14	**25**
03-No knowledgeable household member found after 3 visits	15		2	**17**
04-Temporarily absent/unavailable for field period	42	60	244	**346**
05-No competent household member	16	9	20	**45**
06-Language problem (i.e., between HH member & Survey Team)			3	**3**
07-Dwelling could not be found, insufficient address info			12	**12**
08-Vacant dwelling / selected address has no household		40	202	**242**
09-Dwelling not habitable/destroyed/converted to commercial		8	16	**24**
11-Panel HH has moved; not located	143			**143**
12-Panel HH has united with other panel HH	9			**9**
31-No eligible household member aged 15-64	394	45	219	**658**
32-Refusal by SP to begin modules 2-7	18	3	1	**22**
33-Interview termination-SP unusual circumstance			1	**1**
35-Unable to contact SP after 3 attempts			1	**1**
36-SP absent for entire survey period	1			**1**
40-SP language problem			1	**1**
61-SP unable to read language of General Booklet (GB)	2	1	2	**5**
62-SP refusal to begin GB	7		1	**8**
63-SP unable to do GB due to unusual circumstance			1	**1**
64-SP blindness or visual impairment-unable to do GB	4		2	**6**
65-SP physical disability	5	1	1	**7**
71-SP looked through GB-unable to read/write language of GB	1			**1**
72-SP refused to continue GB after starting	1		3	**4**
74-GB terminated - unusual circumstance	2			**2**
76-Attempted part or all of GB; <3 correct Core answers	5		1	**6**
81-No attempt at any Exercise Booklet (EB) items	22	7	6	**35**
82-Interruption too long during EB administration	1			**1**
83-Unusual circumstance, EB administration terminated			2	**2**
84-Attempt at some EB items, no answers marked	8	3	2	**13**
85-Attempt at all EB items, no answers marked	5	2	5	**12**
86-Marked some EB items, not all items attempted	19	1	8	**28**
87-Marked some EB items, all items attempted	107	19	29	**155**
88-Marked all EB items	1268	288	547	**2103**
95-Household not activated			145	**145**
98-Incorrect eligible household member selected	29	3	5	**37**
99-Ineligible household member selected	1			**1**
Total	**2313**	**633**	**2100**	**5046**

4.2   Final Output Files from Weighting Process

The final output file from the weighting process is,

1) UKR-W01-STEP(HH)- Design-BM Wts (03-Jul-2014)

   a) This file contains the original data variables from the Ukraine data file 'STEP Ukraine - BQ Data for weighting_s7' and the applicable 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 5100 sampled cases.
      i)   2453 participating households with a final household weight,
      ii)   2389 participating selected persons with a final person weight,
      iii) 2389 participating selected persons with a final benchmarked person weight

   c) In addition to the Ukraine 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 6 provides a list of variables from the weighting process that are included in the final weighted data file.

      ii)   Table 7 provides a list of variables from the weighting process that are <u>not</u> included in the final weighted data file. These variables are available from other interim files that were created during the weighting process.

      iii) Table 8 provides the frequency count of result codes for sampled households in the final weighted data file 'UKR-W01-STEP(HH)- Design-BM Wts (03-Jul-2014)' that were activated or should have been activated.

Table 6: Variables Created During Weighting Process - *Added to Ukraine Data File*

Variable Name	Description	Values
Urban_Rural_Ind	Urban Rural Indicator	1-Urban, 2-Rural
PSU_CODE	PSU ID, 2003	n/a
PD_CODE	Postal District ID, 2003	n/a
PDs_Total2012	Postal Districts 2012, Number within PSU	n/a
PDs_Smpd2012	Postal Districts 2012, Number sampled within PSU	n/a
PDs_Total2003E	Postal Districts 2003, Estimated number within PSU	n/a
PDs_Smpd2003E	Postal Districts 2003, Estimated number sampled within PSU	n/a
HHs_Total	HHs, Number of HHs in Selected Postal District	n/a
HHs_PDsplit	Total HHs (estimated) in a PSU_ID	n/a
PSU_Prob2003	PSU Probability, 2003	n/a
SSUProb_Est2003	SSU Estimate of Probability, 2003	n/a
PSUSSU_EstWt	PSU_SSU Estimated 2003 Weight	n/a
HHwt_2007	ULMS 2007 HH Weight	n/a
GATS_VP_No	GATS Voting Precinct #	n/a
GATS_VP_Wgt	GATS Voting Precinct weight	n/a
HHwt_2007	ULMS 2007 Final Household Weight	n/a
HHwt_2007_PSU	Average 2007 PSU HH weight	n/a
RespCat_PSU	Final Response Category-PSUs	1-Participant PSU 3-Non-participant Eligible PSU
RespCat_SSU	Final SSU Response Category	1-Participant PSU_SSU Unit 2-Non-participant PSU_SSU Unit: Eligible
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
W3_FinHHwt	Final Household Weight	n/a
W4_FinSPwt	Final Person Weight	n/a
W3_FinHHwt_C	Comprehensive Household Weight: All Subsamples	n/a
W4_FinSPwt_C	Comprehensive Person Weight: All Subsamples	n/a
BM_SPwt_C	Comprehensive Benchmarked Person Weight, All Subsamples; adjusted for 2013 National Gender & AgeGrp_5yr counts	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'

Table 7:   Variables Created During Weighting Process – Not *Added to Ukraine Data File*

Variable Name	Description	Values
OBLAST_Orig	OBLAST ID #, ULMS 2003 Panel	1 to 26
TYPE_Orig	TYPE of Area, Urban-Rural	1-Urban 2-Rural
OBLAST_Chng	OBLAST Changed Indicator	0-HH in Same OBLAST as Original OBLAST 1-HH Moved to Different OBLAST than Original OBLAST
TYPE_Chng	TYPE Changed Indicator	0-HH in URBAN_RURAL Area as Original TYPE 1-HH Moved to Different URBAN_RURAL Area than Original TYPE
W1_PSU	Basic PSU Weight for Subsample=3	n/a
W1_PSUSSU	Final PSU_SSU Weight for Subsample=2	n/a
W_HH	Basic Household Weight	n/a
W3_HHwt_basic	Basic Household Weight	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_x	PSU - Ineligible SPs	
EP_z	PSU - # of Non-participant SPs in participant HHs	n/a
HH_R_SubPropn	Proportion of Participant HHs in OBLAST Subsample	n/a
SP_R_SubPropn	Proportion of Participant Persons in OBLAST Subsample	n/a
HHsEst_OBLAST	Estimated # of HHs in Oblast	n/a
PopEst_Natl_GenAge	Sample Estimate of Gender-AgeGrp Population	n/a
PopnNatl_2013_GenAge	National 2013 Gender-AgeGrp Urban Population	n/a

Table 8: Ukraine, Frequency/Percent of Final Result Codes For [Activated] Sample Households (excludes final_status=95)

Case Final Result Code	n	% of Total n
01-Refusal-Household	924	18.6%
02-Unusual circumstance-Household	25	.5%
03-No knowledgeable household member found after 3 visits	17	.3%
04-Temporarily absent/unavailable for field period	346	7.0%
05-No competent household member (e.g., severe illness, mental disability, etc.)	45	.9%
06-Language problem (i.e., between HH member & Survey Team)	3	.1%
07-Dwelling could not be found, insufficient Sample Frame address info	12	.2%
08-Vacant dwelling / selected address has no household	242	4.9%
09-Dwelling not habitable/destroyed/converted to commercial	24	.5%
11-Panel HH has moved; not located	141	2.8%
12-Panel HH has united with other panel HH	9	.2%
31-No eligible household member aged 15-64	658	13.3%
32-Refusal by SP to begin modules 2-7	22	.4%
33-Interview termination-SP unusual circumstance	1	.0%
35-Unable to contact SP after 3 attempts	1	.0%
36-SP absent for entire survey period	1	.0%
40-SP language problem	1	.0%
61-SP unable to read language of General Booklet (GB)	5	.1%
62-SP refusal to begin GB	8	.2%
63-SP unable to do GB due to unusual circumstance	1	.0%
64-SP blindness or visual impairment-unable to do GB	6	.1%
65-SP physical disability	7	.1%
71-SP looked through GB-unable to read/write language of GB	1	.0%
72-SP refused to continue GB after starting	4	.1%
74-GB terminated - unusual circumstance	2	.0%
76-Attempted part or all of GB; <3 correct Core answers	6	.1%
81-No attempt at any Exercise Booklet items / refusal to start	35	.7%
82-Interruption too long during Exercise Booklet administration	1	.0%
83-Unusual circumstance, Exercise Booklet administration terminated	2	.0%
84-Attempt at some Exercise Booklet items, no answers marked	13	.3%
85-Attempt at all Exercise Booklet items, no answers marked	12	.2%
86-Marked some Exercise Booklet items, not all items attempted	28	.6%
87-Marked some Exercise Booklet items, all items attempted	155	3.1%
88-Marked all Exercise Booklet items	2103	42.4%
94-HH originally selected in Rural Stratum; data quality issue	56	1.1%
98-Incorrect eligible household member selected	37	.7%
99-Ineligible household member selected	1	.0%
Total	**4955**	**100.0%**

5   <u>Weight Calculations based on World Bank (WB) Weight Requirement</u>

The input file for the process of calculating the weights based on the World Bank (WB) weight requirement is 'W00-1_R-UKR Final STEP URBAN Data File(24-Jun-2014)-for wtg'.

The Ukraine STEP survey included a sample from 3 components. For each STEP component 'C', a final person weight, W_P^C, must be calculated where,

$$1) \quad C = \begin{cases} 1, & ULMS\ Panel\ component \\ 2, & ULMS\ 2012\ component \\ 3, & STEP\ Urban\ Subsample \end{cases}$$

The following sections outline the weighting calculations for each STEP component as well as the weighting calculations.

5.1   ULMS Panel 2003: Basic Household Weight Calculation

For the Ukraine STEP, the ULMS Panel component sample consisted of the participant households from the ULMS-2003 Panel Study and the follow up surveys ULMS-2004 and ULMS-2007. Therefore, the basic household weight for the ULMS Panel component of STEP is the final household weight from the ULMS 2003 Panel and the follow up surveys ULMS-2004 and ULMS-2007; these weights, i.e., w_h, were provided by the Ukraine Survey Team.

For the calculation of w_h , refer to **Appendix I. Description of the Data Weighting in UKRAINIAN LONGITUDINAL MONITORING SURVEY 2003 (ULMS 2003**), which is an excerpt from the document *STEP "Skills Measurement Study + UKRAINIAN LONGITUDINAL MONITORING SURVEY 2012: Technical Report"*.

Note re Household Weight Approximation for ULMS 2003 Panel Households

The Ukraine Survey Team was not able to provide the initial household weight for the ULMS 2003 Panel households, nor was it able to provide the detailed data required for the calculation of the weights for the stages of household selection for the ULMS 2003 Panel households. Instead, the Ukraine Survey Team provided a set of final ULMS 2007 Household Weights in the Excel file 'HH_2007'.

For most ULMS 2003 Panel households, the ULMS 2007 Household Weight (variable 'HHwt_2007' in the weighted data file) was used as an approximation of the final household weight from the ULMS 2003 Panel; for panel households where the ULMS 2007 Household Weight was not available the PSU average ULMS 2007 Household Weight (variable 'HHwt_2007_PSU' in the weighted data file) was used as an approximation of the final household weight from the ULMS 2003 Panel.

For the ULMS Panel component, i.e., C=1, the basic weight for the hth household, $^{C=1}W\_HH_h$, may be represented by the equation,

$$2) \quad ^1W\_HH_h = w_h$$

where,

$$w_h = \begin{cases} HHwt\_2007, & if\ available \\ HHwt\_2007\_PSU, & if\ unavailable\ HHwt\_2007 \end{cases}$$

5.2 ULMS Panel 2012 Subsample: Basic Household Weight Calculation

The ULMS Panel 2012 Subsample of households was selected within each stratum at the third stage of sampling. The first stage unit is the ULMS 2003 Settlement; the second stage unit is a Postal District; the third stage unit is a household; the fourth stage sample unit is an individual aged 15 to 64.

5.2.1 ULMS Panel 2012 Subsample: First Stage Weight (Settlement Weight)

For the ULMS Panel 2012 Subsample component, i.e., C=2, the first stage weight is incorporated in the estimated PSU-SSU weight, say w_j, from the ULMS 2003 Panel Survey.

Note re PSU-SSU Weight Approximation for ULMS 2012 Panel Subsample

The Ukraine Survey Team was not able to provide the actual first stage and second stage weights for the ULMS Panel 2012 Subsample component, nor was it able to provide the detailed data required for the calculation of these weights. The Ukraine Survey Team provided an estimate of the initial PSU-SSU weight (variable 'PSUSSU_EstWt' in the weighted data file) for the ULMS 2012 Panel Subsample; this estimated weight was based on data from the ULMS 2003 Panel and the ULMS 2012 Panel Subsample.

The estimated initial PSU-SSU weight is provided in the Excel file 'STEP_panel_part_with 2003 prob_1 - Revd 24-Jun-2014-1'.

5.2.2 ULMS Panel 2012 Subsample: Second Stage Weight (Postal District Weight)

For STEP component C=2, the basic second stage weight for the jth PSU-SSU unit in stratum 'g', $^2W_{gj}$, is represented by the equation:

$$3) \qquad {}^2W_{gj} = w_{gj} \times \frac{d}{d'} = w_{gj} \times \frac{280}{115} \approx w_{gj} \times 2.4348$$

where,

w_{gj} is the estimated PSU-SSU weight (variable 'PSUSSU_EstWt' in the weighted data file) based on data from the ULMS 2003 Panel and the ULMS 2012 Panel subsample; these estimated PSU-SSU weights were provided by the Ukraine Survey Team;

$d=280$ is the number of urban PSU-SSU units selected in the ULMS 2003 Panel;

$d'=115$ is the number of urban PSU-SSU units sub-sampled from the ULMS 2003 Panel urban PSU-SSU units.

N.B.   the values d=280 & d'=115 were deduced from the Excel file 'STEP_panel_part_with 2003 prob_1 -Revd 24-Jun-2014-1'.

➢ **The basic second stage estimated PSU-SSU weight is appended to the STEP weighted data file.**

5.2.3 ULMS Panel 2012 Subsample: Second-stage Response

A second-stage response adjustment was carried out for the PSU-SSUs within the ULMS Panel 2012 Subsample component.

Regarding the ULMS Panel 2012 Subsample component, for treatment in the weighting process the data collection outcomes for the sampled SSUs may be summarized in two possible categories: participant, non-participant with known eligibility.

Participant - A participating SSU is an <u>eligible</u> sampled SSU that provides sufficient data to be considered a participant in the survey. The sampled SSUs with one or more person-level result codes 32-40, 51-88 or 98-99 would satisfy the requirement for classification of a SSU as a survey participant.

Non-participant, Eligibility known:

There was one non-participating SSU, SSU # 272 in Oblast 16, in the ULMS Panel 2012 Subsample. Although all 5 sampled households in SSU # 272 were categorized as ineligible (i.e., data variable *'m11_q07, Result code of interview'=31, No household member in the eligible range of 15-64*), this SSU is treated in the weighting process as an eligible SSU <u>since SSU # 272 is a participant SSU in the ULMS 2003 Panel component of the Ukraine STEP Survey</u>. Thus, it is clear that SSU # 272 included some eligible households, and is treated as an eligible SSU rather than an ineligible SSU in the weighting process.

For the ULMS Panel 2012 Subsample component, the total number of sampled SSUs may be represented by the following equation.

4) $y_j = y_k + y_z$

where,

    y_j - the number of sampled SSUs (i.e., y_j = 115 urban SSUs),
    y_k - the number of participant SSUs 'k',
    y_z - the number of non-participating SSUs 'z' with known eligibility.

5.2.3.1 ULMS Panel 2012Subsample: Response Adjustment Factor for SSU Non-participation

An SSU response adjustment is necessary to compensate for SSUs that did not participate in the survey. The SSU response adjustment factor for the jth SSU, R_j, is calculated as follows:

5) $^{2}\boldsymbol{R_j} = \begin{cases} \frac{Total\ eligible\ SSUs}{Total\ particpating\ SSUs} = \frac{y_k + y_z}{y_k}, & Participant\ SSU \\ 0, & Non\text{-}participant\ SSU \end{cases}$

5.2.4   ULMS Panel 2012 Subsample: Final SSU Weight

The final SSU weight, i.e., the basic response-adjusted SSU weight, is calculated as the product of the basic PSU weight and the SSU response adjustment factor. Therefore, for SSU 'j', the final SSU weight, $^{2}W2_{gj, \text{ is}}$ calculated as follows:

6) $^{2}W2_{gj} = \ ^{2}W_{gj} * \ ^{2}R_{j}$

➢ **The final SSU weight is appended to the STEP data file.**

5.2.5   ULMS Panel 2012 Subsample: Third-stage Weight (Household Weight)

For the ULMS Panel 2012 Subsample component, the initial third stage weight for the hth household in Postal District 'j', $^{C=2}W3_{jh}$, is calculated as follows:

7) $^{2}W0_{jh} = \frac{\hat{N}_{j}}{n_{j}}$

where,

$\hat{N}_{j}$ -   the estimated number of <u>listed households</u> (i.e., data file variable '*HHs_PDsplit*') in the jth SSU,

n_{j} -   the number of households sampled in the jth SSU.

The basic third stage weight is similarly calculated for each selected Postal District in Subsample 2.

5.2.6 ULMS Panel 2012 Subsample: Basic Household Weight

For the ULMS Panel 2012 component, i.e., C=2, the Basic Household weight is the product of the weights from the first three sampling stages. Therefore, the Basic Household Weight for the ith household that incorporates the PSU weight, the Postal District weight, and the Household weight is calculated as follows:

8) $\quad {}^{2}W\_HH_h = {}^{2}W2_{gj} * {}^{2}W0_{jh}$

Example 5:   **Basic Household Weight**

In Subsample 2, the third stage weight for each of the 6 sampled households in PSU_ID # 42, , which contains an estimated 8422 households is calculated as follows:

$${}^{2}W\_HH_h = {}^{2}W2_{gj} * {}^{2}W0_{jh} = 19.2158 * 1403.67 \approx 26972.5191$$

5.3 STEP Urban Subsample of Voting Precincts: Basic Household Weight Calculation

The STEP Urban Subsample of households was selected within each stratum at the second stage of sampling; the first stage unit is the Voting Precinct, the second stage unit is a household. Subsequently, an individual aged 15 to 64 was selected at the third stage of sampling.

5.3.1 STEP Urban Subsample of Voting Precincts: First Stage Weight (PSU Weight)

The STEP Urban Subsample component's primary sample unit (PSU) was a Voting Precinct; 70 Voting Precincts were systematically selected from the 300 Voting Precincts that had been selected for the Ukraine Global Adult Tobacco Survey (GATS), 2010.

For the STEP Urban Subsample component, i.e., C=3, the basic first stage Voting Precinct weight the jth Voting Precinct, $^{C=3}W1_j$, is calculated as follows:

$$9) \quad {}^{3}W_j = W_{VP,GATS} * \frac{V}{v}$$

where,

V = 300,	the number of GATS sampled Voting Precincts;
v = 70,	the number of STEP Voting Precincts selected from the 300 GATS sample Voting Precincts;
W$_{VP, GATS}$ -	the GATS Voting Precinct Weight provided by the Ukraine Survey Team.

➢ **The basic STEP first stage Voting Precinct weight is appended to the STEP weighted data file.**

Example 6:   **First Stage Weight**

The basic first stage weight for PSU # 1 in Subsample 3:

$$^{3}W_j = W_{VP,GATS} * \frac{V}{v} = 43.4638 * \frac{300}{70} \approx 186.2736$$

The basic first stage weight is similarly calculated for each of the 70 sampled urban PSUs.

5.3.2 STEP Urban Subsample of Voting Precincts: First-stage Response Adjustment

A first-stage response adjustment was carried out for the PSUs within the STEP Urban Subsample of Voting Precincts component of the Ukraine Step Survey.

Regarding the STEP Urban Subsample of Voting Precincts component, for treatment in the weighting process the data collection outcomes for the sampled PSUs may be summarized in two possible categories: participant, non-participant with known eligibility.

Participant - A participating PSU is an <u>eligible</u> 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 or 98-99 would satisfy the requirement for classification of a PSU as a survey participant.

Non-participant - Eligibility known: In the STEP Urban Subsample of Voting Precincts component of the Ukraine Step Survey, a non-participating PSU with known eligibility resulted when all households in a sampled PSU were not contacted (i.e., data variable *'m11_q07, Result code of interview'=99, Not contacted*). The PSUs 8, 9, and 62 were not contacted for interviewing; however, <u>it is assumed that each of these PSUs included some eligible households</u>.

For the STEP Urban Subsample of Voting Precincts component, the total number of sampled PSUs may be represented by the following equation.

$$10) \quad y_j = y_k + y_z$$

where,

y_j -   the number of sampled PSUs,
y_k -   the number of participant PSUs 'k',
y_z -   the number of non-participating PSUs 'z' with known eligibility.

Note: $y_j = 70$ PSUs in the STEP Urban Subsample of Voting Precincts.

5.3.2.1 STEP Urban Subsample of Voting Precincts: Response Adjustment Factor for PSU 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 jth PSU in Stratum 'g', $^3R1_{gj}$, is calculated as follows:

$$11) \quad {}^3R1_{gj} = \begin{cases} \frac{Total\ eligible\ PSUs\ in\ Stratum\ 'g'}{Total\ particpating\ PSUs\ in\ Stratum\ 'g'} = \frac{y_{gk}+\ y_{gz}}{y_{gk}}, & Participant\ PSU \\ 0, & Non\text{-}participant\ PSU \end{cases}$$

Example 7:   Response-adjustment Factor

For the STEP Urban Subsample of Voting Precincts component, PSU response adjustment factor is calculated as follows:

$$^3R1_{gj} = \frac{y_k +\ y_z}{y_k} = \frac{67 + 3}{67} \approx\ 1.0448$$

5.3.3 STEP Urban Subsample of Voting Precincts: Final PSU Weight

The final PSU weight, i.e., the basic 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 final PSU weight, $W1_{j, is}$ calculated as follows:

$$12) \quad {}^{3}W1_j = {}^{3}W_j * R1_j$$

> ➢ **The final PSU weight is appended to the STEP data file.**

Example 8:   <u>**Final PSU Weight, i.e., Basic Response-adjusted PSU Weight**</u>

PSU # 1 in Subsample 3,

$$\begin{aligned} {}^{3}W1_j &= {}^{3}W_j * {}^{3}R1_j \\ &\approx 186.2736 \times 1.0448 \approx 194.6142 \end{aligned}$$

The final PSU weight is similarly calculated for all sampled PSUs in the STEP Urban Subsample of Voting Precincts component.

5.3.4 STEP Urban Subsample of Voting Precincts: Second-stage Weight (Household Weight)

For the STEP Urban Subsample component, i.e., C=3, the initial second stage weight for the hth household in Voting Precinct 'j', ${}^{3}W_{jh}$, is calculated as follows:

$$13) \quad {}^{3}W0_{jh} = \frac{N_j}{n_j}$$

where,

     N_j - the number of <u>listed households</u> in the jth Voting Precinct,

     n_j - the number of households sampled in the jth Voting Precinct; for C=3, n_j = 30, which includes the initial sample households plus the reserve sample households.

Example 9:   <u>**Initial Household Weight**</u>

For PSU # 1 with 723 households in Subsample 3,

$$ {}^{3}W0_{jh} = \frac{N_j}{n_j} = \frac{723}{30} = 24.1 $$

The basic second stage weight is similarly calculated for each of the selected Voting Precincts.

5.3.5 STEP Urban Subsample of Voting Precincts: Basic Household Weight

For the STEP Urban Subsample component, i.e., C=3, the Basic Household weight for the hth household is the product of the weights from the first two sampling stages. Therefore, the Basic Household Weight that incorporates the Voting Precinct weight and the Household weight is calculated as follows:

$$14) \quad {}^{3}W\_HH_h = {}^{3}W1_j * {}^{3}W0_{jh}$$

Example 10:   Basic Household Weight

For PSU # 1 with 723 households in Subsample 3,

$$^3W\_HH_h = \ ^3W1_j *\ ^3W0_{jh} = 194.6142 * 24.1 \approx 4690.20$$

5.4 Response Adjustment for Basic Household Weights (All Components)

The response adjustment was carried out <u>for each STEP component</u>; the response adjustment procedure was the same for each Step component.

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

Participant - A participating household '**k**' is an <u>eligible</u> sampled household that provides sufficient data to be considered a participant in the survey. The sampled households '**k**' with person-level result codes 32-40, 51-88 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 <u>ineligible</u> if there is no household member in the eligible 15-64 age range. The result codes for ineligible sampled households '**x**' are as follows:

II. Household-level Codes – Ineligible Household	
Code	**Code Description**
07	Dwelling could not be found / selected address has no household
08	Vacant dwelling
09	Dwelling not habitable/dwelling destroyed/dwelling converted to commercial use
12	Panel HH has united with other panel HH
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**
11	Panel HH has moved; not located
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.

$$15) \quad n_j = n_{jk} + n_{jx} + n_{jz} + n_{ju}$$

where,

n_j - the number of households sampled in the jth PSU,

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

n_{jx} - the number of ineligible households 'x' in the jth PSU,

n_{jz} - the number of non-participating households 'z' with known eligibility in the jth PSU,

n_{ju} - the number of non-participating households 'u' with unknown eligibility in the jth PSU.

Tables 9-11 provide, for Subsamples 1-3, the counts of sampled households from the Ukraine data file by the outcome categories described above.

Table 9 provides, for Subsample 1, i.e., the ULMS 2003 Panel component of STEP, the counts of sampled households from the Ukraine data file by the outcome categories described above.

Table 9: Subsample=1, Number of Sampled Households by Final Response Category						
	Final Response Category-Households					
Case Final Result Code	Participant (n_{jk})	Ineligible (n_{jx})	Non-participant Eligible (n_{jz})	Non-participant Eligibility Unknown (n_{ju})	Total (n_j)	% of Total
01-Refusal-Household				177	177	7.7%
02-Unusual circumstance-HH				11	11	.5%
03-No knowledgeable HH member found				15	15	.6%
04-Temporarily absent				42	42	1.8%
05-No competent HH member				16	16	.7%
11-Panel HH has moved; not located			143		143	6.2%
12-Panel HH has united with other panel HH		9			9	.4%
31-No eligible HH member aged 15-64		394			394	17.0%
32-Refusal by SP to begin modules 2-7	18				18	.8%
36-SP absent for entire survey period	1				1	.0%
61-SP unable to read General Booklet (GB) language	2				2	.1%
62-SP refusal to begin GB	7				7	.3%
64-SP visual impairment-unable to do GB	4				4	.2%
65-SP physical disability	5				5	.2%
71-SP unable to read/write language of GB	1				1	.0%
72-SP refused to continue GB	1				1	.0%
75-Attempted GB, no answers marked	2				2	.1%
76-Attempted GB; <3 correct Core answers	5				5	.2%
81-No attempt at Exercise Booklet(EB) items	22				22	1.0%
82-Interruption too long-EB termination	1				1	.0%
84-Attempt at some EB items, no answers marked	8				8	.3%
85-Attempt at all EB items, no answers marked	5				5	.2%
86-Marked some EB items, some items attempted	19				19	.8%
87-Marked some EB items, all items attempted	107				107	4.6%
88-Marked all EB items	1268				1268	54.8%
98-Incorrect eligible HH member selected	29				29	1.3%
99-Ineligible household member selected	1				1	.0%
Total	1506	403	143	261	2313	100.0%

Table 10 provides, for Subsample 2, i.e., the ULMS 2012 Subsample component of STEP, the counts of sampled households from the Ukraine data file by the outcome categories described above.

Table 10: Subsample=2, Number of Sampled Households by Final Response Category						
	Final Response Category-Households					
Case Final Result Code	Participant (n_{jk})	Ineligible (n_{jx})	Non-participant Eligible (n_{jz})	Non-participant Eligibility Unknown (n_{ju})	Total (n_j)	% of Total
01-Refusal-Household				143	143	22.8%
04-Temporarily absent				60	60	9.6%
05-No competent HH member				9	9	1.4%
08-Vacant dwelling		40			40	6.4%
09-Dwelling not habitable		8			8	1.3%
31-No eligible HH member aged 15-64		40			40	6.4%
32-Refusal by SP to begin modules 2-7	3				3	.5%
61-SP unable to read General Booklet (GB) language	1				1	.2%
65-SP physical disability	1				1	.2%
81-No attempt at Exercise Booklet(EB) items	7				7	1.1%
84-Attempt at some EB items, no answers marked	3				3	.5%
85-Attempt at all EB items, no answers marked	2				2	.3%
86-Marked some EB items, some items attempted	1				1	.2%
87-Marked some EB items, all items attempted	19				19	3.0%
88-Marked all EB items	288				288	45.9%
98-Incorrect eligible HH member selected	3				3	.5%
Total	328	88	0	212	628	100.0%

Table 11 provides, for Subsample 3, i.e., the STEP Urban Subsample of Voting Precincts, the counts of sampled households from the Ukraine data file by the outcome categories described above.

Table 11: Subsample=3, Number of Sampled Households by Final Response Category						
	Final Response Category-Households					
Case Final Result Code	Participant (n_{jk})	Ineligible (n_{jx})	Non-participant Eligible (n_{jz})	Non-participant Eligibility Unknown (n_{ju})	Total (n_j)	% of Total
01-Refusal-Household				604	604	30.9%
02-Unusual circumstance-Household				14	14	.7%
03-No knowledgeable household member found				2	2	.1%
04-Temporarily absent				244	244	12.5%
05-No competent HH member				20	20	1.0%
06-Language problem (HH contact member)				3	3	.2%
07-Dwelling could not be found, insufficient address				12	12	.6%
08-Vacant dwelling		202			202	10.3%
09-Dwelling not habitable		16			16	.8%
31-No eligible HH member aged 15-64		219			219	11.2%
32-Refusal by SP to begin modules 2-7	1				1	.1%
33-Interview termination-SP unusual circumstance	1				1	.1%
35-Unable to contact SP after 3 attempts	1				1	.1%
40-SP language problem	1				1	.1%
61-SP unable to read General Booklet (GB) language	2				2	.1%
62-SP refusal to begin GB	1				1	.1%
63-SP unable to do GB due to unusual circumstance	1				1	.1%
64-SP visual impairment-unable to do GB	2				2	.1%
65-SP physical disability	1				1	.1%
72-SP refused to continue GB	3				3	.2%
76-Attempted GB; <3 correct Core answers	1				1	.1%
81-No attempt at Exercise Booklet(EB) items	6				6	.3%
83-Unusual circumstance, EB terminated	2				2	.1%
84-Attempt at some EB items, no answers marked	2				2	.1%
85-Attempt at all EB items, no answers marked	5				5	.3%
86-Marked some EB items, some items attempted	8				8	.4%
87-Marked some EB items, all items attempted	29				29	1.5%
88-Marked all EB items	547				547	28.0%
98-Incorrect eligible HH member selected	5				5	.3%
Total	619	437	0	899	1955	100.0%

5.4.1 Eligible Proportion of Households

The STEP Survey assumes that the households 'u' with unknown eligibility in the jth 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

16) $\quad {}^c\varepsilon_j = \dfrac{Total\ \#\ of\ known\ eligible\ cases}{Total\ \#\ of\ known\ eligible\ \&\ known\ ineligible\ cases} = \dfrac{n_{jk} + n_{jz}}{n_{jk} + n_{jx} + n_{jz}}$

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

Example 11:   Eligible Proportion of Households

Subsample=1, for PSU_CODE #101, the following household results were obtained:

n_j	n_{jk}	n_{jx}	n_{jz}	n_{ju}
37	30	4	1	2

For PSU_CODE #101 in Subsample 1, the proportion of households with known eligibility is calculated as,

$${}^1\varepsilon_{101} = \dfrac{Total\ \#\ of\ known\ eligible\ cases}{Total\ \#\ of\ known\ eligible\ \&\ known\ ineligible\ cases}$$

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

$$= \dfrac{30 + 1}{30 + 4 + 1} = \dfrac{31}{35} \approx 0.8857$$

Example 12:   Eligible Proportion of Households

Subsample=2, for PSU_CODE # 301, the following household results were obtained:

n_j	n_{jk}	n_{jx}	n_{jz}	n_{ju}
6	3	2	0	1

For PSU_CODE # 301 in Subsample 2, the proportion of households with known eligibility is calculated as,

$${}^1\varepsilon_{301} = \dfrac{Total\ \#\ of\ known\ eligible\ cases}{Total\ \#\ of\ known\ eligible\ \&\ known\ ineligible\ cases}$$

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

$$= \dfrac{3 + 0}{3 + 2 + 0} = \dfrac{3}{5} \approx 0.60$$

Example 13:   <u>Eligible Proportion of Households</u>

Subsample=3, for PSU_CODE # 101, the following household results were obtained:

n_j	n_{jk}	n_{jx}	n_{jz}	n_{ju}
30	4	10	0	16

For PSU_CODE # 101in Subsample 3, the proportion of households with known eligibility is calculated as,

$$^1\varepsilon_{101} = \frac{Total\ \#\ of\ known\ eligible\ cases}{Total\ \#\ of\ known\ eligible\ \&\ known\ ineligible\ cases}$$

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

$$= \frac{4 + 0}{4 + 10 + 0} = \frac{4}{14} \approx 0.286$$

5.4.2 Response Adjustment Factor for Non-participation

For each STEP component, 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 jth PSU in Component 'C', $^{C}R_j$, is calculated as follows:

$$17) \quad ^{C}R_j = \frac{Total\ eligible\ households\ in\ PSU\ 'j'}{Total\ particpating\ households\ in\ PSU\ 'j'} = \frac{n_{jk} + n_{jz} + (\varepsilon_j * n_{ju})}{n_{jk}}$$

➢ The response adjustment ratio, $^{C}R_j$, 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 14:   Response Adjustment Factor for sampled Households

Subsample=1, for PSU_CODE #101, the following household results were obtained:

n_j	n_{jk}	n_{jx}	n_{jz}	n_{ju}
37	30	4	1	2

Therefore,

$$^{1}R_{101} = \frac{Total\ eligible\ households\ in\ PSU\ \#101}{Total\ particpating\ households\ in\ PSU\ \#101}$$

$$= \frac{n_{jk} + n_{jz} + (\varepsilon_{101} * n_{ju})}{n_{jk}}$$

$$= \frac{30 + 1 + (.8857 * 2)}{30} \approx 1.0924$$

The response adjustment ratio, $^{C}R_j$, is similarly calculated for all sampled PSUs.

Example 15:   Response Adjustment Factor for sampled Households

Subsample=2, for PSU_CODE # 301, the following household results were obtained:

n_j	n_{jk}	n_{jx}	n_{jz}	n_{ju}
6	3	2	0	1

Therefore,

$$^{2}R_{301} = \frac{Total\ eligible\ households\ in\ PSU\ \#201}{Total\ particpating\ households\ in\ PSU\ \#201}$$

$$= \frac{n_{jk} + n_{jz} + (\varepsilon_{101} * n_{ju})}{n_{jk}}$$

$$= \frac{3 + 0 + (.6 * 1)}{3} \approx 1.20$$

Example 16: <u>**Response Adjustment Factor for sampled Households**</u>

Subsample=3, for PSU_CODE # 101, the following household results were obtained:

n_j	n_{jk}	n_{jx}	n_{jz}	n_{ju}
30	4	10	0	16

Therefore,

$$^3R_{101} = \frac{Total\ eligible\ households\ in\ PSU\ \#101}{Total\ particpating\ households\ in\ PSU\ \#101}$$

$$= \frac{n_{jk} + n_{jz} + \left(\varepsilon_{101} * n_{ju}\right)}{n_{jk}}$$

$$= \frac{4 + 0 + (0.286 * 16)}{4} \approx 2.1429$$

5.5 Subsample Household Weight (Adjusted for Non-response)

The response-adjusted Subsample household weight is calculated as the product of the basic household weight and the response adjustment factor for non-participant households. Therefore, for household 'h', the Subsample response-adjusted household weight, $_R^C W_h$, (i.e., final data file variable 'W3_FinHHwt') is calculated as follows:

$$18) \quad _R^C W_h = {}^C W\_HH_h * {}^C R_j$$

➢ **The Subsample response-adjusted household weight is appended to the STEP data file.**

Example 17:   **Response-adjusted Subsample Household Weight**

In Subsample 1, for PSU # 1, the subsample household weight for participating households is calculated as follows:

$$_R^1 W_h = {}^1 W\_HH_h * {}^1 R_j = 4957.8226 * 1.0924$$
$$= 5415.8309$$

Example 18:   **Response-adjusted Subsample Household Weight**

In Subsample 2, for PSU_ID # 42, the subsample household weight for each of the three participating households is calculated as follows:

$$_R^2 W_h = {}^2 W\_HH_h * {}^2 R_j = 26972.5191 * 1.20$$
$$= 32367.0229$$

Example 19:   **Response-adjusted Subsample Household Weight**

In Subsample 3, for PSU # 1, the subsample household weight for participating households is calculated as follows:

$$_R^3 W\_HH_h = {}^3 W\_HH_h * {}^3 R_j = 4690.20 * 2.1429$$
$$= 10050.4336$$

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

5.6 Subsample Proportion Adjustment for Household Weight

The Ukraine STEP sample is comprised of three subsample components:
1) ULMS Panel Sample, i.e., subsample 1, C=1;
2) New ULMS-2012 subsample i.e., subsample 2, C=2;
3) STEP Voting Precincts Urban Subsample i.e., subsample 3, C=3.

The participant household sample from each STEP subsample component comprises the overall STEP final sample of participating households. Thus, the final sample of participating households from each STEP subsample component is a proportion of the overall STEP final sample of participant households.

Each component subsample's proportionate share of the overall STEP <u>household</u> sample is hereafter referred to as the Household Subsample Proportionate Adjustment Factor, $^{C}T_g$.

The Household Subsample Proportionate Adjustment Factor is calculated <u>at the stratum level</u> as follows:

$$19) \quad {}^{C}T_g = \frac{{}^{c}n_g}{{}^{1}n_g + {}^{2}n_g + {}^{3}n_g} = \frac{{}^{c}n_g}{n_g}$$

where, $^{c}n_g$ is the number of participant households in subsample C in stratum 'g',

n_g is the total number of participant households in stratum 'g',

i.e., $n_g = {}^{1}n_g + {}^{2}n_g + {}^{3}n_g$

Example 20:   Household Subsample Proportionate Adjustment Factor

The Household Subsample Proportionate Adjustment Factors in stratum 1 which has 112 participating households is calculated as follows:

For Subsample 1, in stratum 1, ${}^{1}T_1 = \frac{n^1}{n^1 + n^2 + n^3} = \frac{73}{73+17+22} = \frac{73}{112} \approx 0.6518$

For Subsample 2, in stratum 1, ${}^{2}T_1 = \frac{n^1}{n^1 + n^2 + n^3} = \frac{17}{73+17+22} = \frac{17}{112} \approx 0.1518$

For Subsample 3, in stratum 1, ${}^{3}T_1 = \frac{n^1}{n^1 + n^2 + n^3} = \frac{22}{73+17+22} = \frac{22}{112} \approx 0.1964$

Table 12 provides the Household Subsample Proportionate Adjustment Factor, $^{C}T_g$, by Stratum, i.e., by OBLAST.

Table 12:	Case Summaries: Subsample Number of Participant HHs & Proportion of Participant HHs By OBLAST							
	Subsample *				**% of OBLAST Participant HHs ($^{C}T_g$)**			
OBLAST	**1**	**2**	**3**	**Total**	**1**	**2**	**3**	**Total**
1	73	17	22	112	65.2%	15.2%	19.6%	**100.0%**
2	89	27	17	133	66.9%	20.3%	12.8%	**100.0%**
3	21	9	23	53	39.6%	17.0%	43.4%	**100.0%**
4	24	9	30	63	38.1%	14.3%	47.6%	**100.0%**
5	14	6	12	32	43.8%	18.8%	37.5%	**100.0%**
6	111	27	51	189	58.7%	14.3%	27.0%	**100.0%**
7	262	48	100	410	63.9%	11.7%	24.4%	**100.0%**
8	25	6	27	58	43.1%	10.3%	46.6%	**100.0%**
9	5	3	9	17	29.4%	17.6%	52.9%	**100.0%**
10	83	15	21	119	69.7%	12.6%	17.6%	**100.0%**
11	43	12	12	67	64.2%	17.9%	17.9%	**100.0%**
12	42	6	15	63	66.7%	9.5%	23.8%	**100.0%**
13	129	27	30	186	69.4%	14.5%	16.1%	**100.0%**
14	56	12	22	90	62.2%	13.3%	24.4%	**100.0%**
15	35	9	11	55	63.6%	16.4%	20.0%	**100.0%**
16	60	12	39	111	54.1%	10.8%	35.1%	**100.0%**
17	48	12	30	90	53.3%	13.3%	33.3%	**100.0%**
18	31	3	11	45	68.9%	6.7%	24.4%	**100.0%**
19	42	8	30	80	52.5%	10.0%	37.5%	**100.0%**
20	7	6	13	26	26.9%	23.1%	50.0%	**100.0%**
21	139	23	23	185	75.1%	12.4%	12.4%	**100.0%**
22	40	9	15	64	62.5%	14.1%	23.4%	**100.0%**
23	22	5	13	40	55.0%	12.5%	32.5%	**100.0%**
24	42	5	13	60	70.0%	8.3%	21.7%	**100.0%**
25	17		11	28	60.7%	0.0%	39.3%	**100.0%**
26	46	12	19	77	59.7%	15.6%	24.7%	**100.0%**
Total	**1506**	**328**	**619**	**2453**	**61.4%**	**13.4%**	**25.2%**	**100.0%**

* **1-ULMS-2003 Panel**
2-New ULMS-2012
3-STEP urban Subsample

5.7 Final Household Weight (All Components)

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

The final household weight incorporates the response-adjusted subsample household weight and its proportionate share of the overall STEP sample of participant households. The final comprehensive household weight is determined by adjusting the final subsample household weight by its Household Subsample Proportionate Adjustment factor, $^{C}T_g$,

Therefore, for household 'h' in stratum 'g', the final household weight, $^{C}W_{gh}$, is calculated as follows:

$$20) \quad {}^{c}W_{gh} = {}^{C}_{R}W_h * {}^{C}T_g$$

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

Example 21:   Final Household Weight

i)   In stratum 1, the final household weight for selected HH_ID # 614 in subsample 1 is,

$$^{1}W_{1,614} = {}^{1}_{R}W_{614} * {}^{1}T_1 = 9806.6866 * 0.6519 \approx 6391.8582$$

ii)   In stratum 1, the final household weight for selected HH_ID # 13736 in subsample 2 is,

$$^{3}W_{1,13736} = {}^{3}_{R}W_{13736} * {}^{3}S_1 = 27599.4061 * 0.1518 \approx 4189.1956$$

iii)   In stratum 1, the final household weight for selected HH_ID # 50114 in subsample 3 is,

$$^{3}W_{1,50114} = {}^{3}_{R}W_{50114} * {}^{3}S_1 = 10050.4336 * 0.1964 \approx 1974.1923$$

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.

5.8 Basic Person Weight

<u>For each STEP subsample component</u>, the basic weight for the ith person selected in the hth household, W_{hi} , is calculated as follows:

$$21)\quad W_{hi} = \frac{a_h}{b_h} = \frac{a_h}{1} = a_h$$

where,

a_h - the number of eligible persons in the hth household,

b_h- the number of persons sampled in the hth household, (i.e., n_h =1 person in each Ukraine household that was sampled for STEP).

N.B. $A_{jh} = \sum_{h=1}^{n_{jh}} a_{jh} = total~\#~of~eligible~persons~in~the~j^{th}~PSU$

where,

n_{jh} - the number of households 'h' that participate in the jth PSU

Example 22: **Basic Person Weight within sampled Household**

Selected HH_ID # 401 has 3 eligible persons. Therefore,

$$W_{hi} = \frac{a_{jh}}{b_{jh}} = \frac{3}{1} = 3$$

5.8.1 Person-level Response Adjustment

For eligible participating households in each STEP subsample component, 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 '**k**' is an eligible sampled person that provides sufficient data to be considered a participant in the survey. The sampled persons '**k**' with result codes 51-88 satisfy the requirement for classification of the person as a survey participant. The result codes for participating persons '**k**' 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:

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

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

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

The person weight is only relevant for sampled households in the final response category RespCat_HH=1, i.e., participant households. Table 13 provides by subsample component the pertinent frequency counts of the Final Status Codes for the participant households.

Table 13:   Final Selected Person Status Results for Participant Households

Case Final Result Code	1-ULMS-2003 panel	2-New ULMS-2012	3-STEP urban Subsample	Total	% of Total n Total
32-Refusal by SP to begin modules 2-7	18	3	1	22	.9%
33-Interview termination-SP unusual circumstance			1	1	.0%
35-Unable to contact SP after 3 attempts			1	1	.0%
36-SP absent for entire survey period	1			1	.0%
40-SP language problem			1	1	.0%
61-SP unable to read language of General Booklet (GB)	2	1	2	5	.2%
62-SP refusal to begin GB	7		1	8	.3%
63-SP unable to do GB due to unusual circumstance			1	1	.0%
64-SP blindness or visual impairment-unable to do GB	4		2	6	.2%
65-SP physical disability	5	1	1	7	.3%
71-SP looked through GB-unable to read/write language of GB	1			1	.0%
72-SP refused to continue GB after starting	1		3	4	.2%
74-GB terminated - unusual circumstance	2			2	.1%
76-Attempted part or all of GB; <3 correct Core answers	5		1	6	.2%
81-No attempt at any Exercise Booklet (EB) items	22	7	6	35	1.4%
82-Interruption too long during EB administration	1			1	.0%
83-Unusual circumstance, EB administration terminated			2	2	.1%
84-Attempt at some EB items, no answers marked	8	3	2	13	.5%
85-Attempt at all EB items, no answers marked	5	2	5	12	.5%
86-Marked some EB items, not all items attempted	19	1	8	28	1.1%
87-Marked some EB items, all items attempted	107	19	29	155	6.3%
88-Marked all EB items	1268	288	547	2103	85.7%
98-Incorrect eligible household member selected	29	3	5	37	1.5%
99-Ineligible household member selected	1			1	.0%
Total	**1506**	**328**	**619**	**2453**	**100.0%**

In each participating household in the STEP Survey, one eligible person was selected to be interviewed. Therefore, for each participating household the basic person weight, W_{jis} , 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:

$$22) \quad A_{jh} = A_{jhk} + A_{jhq}$$

where,

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

$A_{jhk} = \sum_{h=1}^{n_{jh}} a_{jhk}$ - the total number of participant eligible persons 'k' in the jth PSU,

$A_{jhq} = \sum_{h=1}^{n_{jh}} a_{jhq}$ - the total number of non-participating eligible persons 'q' in the jth PSU,

Therefore, the person-level response adjustment for the jth PSU, R_SP$_j$, is calculated as follows:

$$23) \quad R\_SP_j = \frac{Total\ eligible\ persons\ in\ all\ participating\ sampled\ households\ in\ PSU\ 'j'}{Total\ participant\ persons\ in\ all\ participating\ sampled\ households\ in\ PSU\ 'j'} = \frac{A_{jh}}{A_{jhk}}$$

Example 23:   **Basic Person Weight Response Adjustment**

i) For PSU #103 in Subsample 1, there are 33 eligible persons across all participating households; also, all selected persons participated. Therefore,

$A_{jh} = \sum_{h=1}^{n_{jh}} a_{jh} = 33$; $A_{jhk} = \sum_{h=1}^{n_{jh}} a_{jhk} = 33$

$R\_SP_j = \dfrac{A_{jh}}{A_{jhk}} = \dfrac{33}{33} = 1.00$

ii) For PSU # 101, the total number of eligible persons across all participating households is 63. For PSU # 101, there are 2 non-participating selected persons Therefore,

$A_{jh} = \sum_{h=1}^{n_{jh}} a_{jh} = 63$; $A_{jh} = \sum_{h=1}^{n_{jh}} a_{jh} = 61$

$R\_SP_j = \dfrac{A_{jh}}{A_{jhk}} = \dfrac{63}{61} \approx 1.03279$

Table 14 provides the sample counts of the Final Response Category for selected persons.

Table 14: Sample Counts by Selected Person Final Response Category for Participating HHs

Final Response Category: Selected Person	Number of Participant HHs (n)				% of Total n
	1-ULMS-2003 Panel	2-New ULMS-2012	3-STEP Urban Subsample	Total	Total
1-Participant SP	1457	322	610	2389	97.4%
2-Non-participant Eligible SP	48	6	9	63	2.6%
3-Ineligible SP	1			1	0.0%
Total	**1506**	**328**	**619**	**2453**	**100.0%**

Based on the above table, there are 2453 participating selected persons for which a final non-zero Person-weight was calculated. The sampled persons in categories 2 and 3 have a zero-value final person weight.

5.8.2 Subsample Person Weight (Adjusted for Non-response)

The response-adjusted subsample person weight is calculated as the product of the response-adjusted subsample household weight, the basic subsample person weight, and the response adjustment for the basic subsample person weight. Therefore, for the selected person 'i' in household 'k' in PSU 'j', the response-adjusted subsample person weight, $^C W_i$ (i.e., weighted data file variable '**W4_FinSPwt**'), is calculated as follows:

$$24) \quad {}_R^C W_i = {}_R^C W_h * W_{hi} * R\_SP_j$$

> **The final response-adjusted person weight is appended to the STEP data file.**

Example 24:   **Response-adjusted Subsample Person Weight**

i)   In PSU #103, selected HH_ID # 614 with 2 eligible persons, the selected person was a participant. All selected persons participated in PSU #103, therefore, $R\_SP_j = 1$

$${}_R^1 W_{614} \approx 9806.6866 \,; \qquad W_{hi} = 2 \,; \qquad R\_SP_j = 1$$

$${}_R^1 W_i = 9806.6866 * 2 * 1 \approx 19613.3732$$

ii)   In PSU #101, in participant household # 401 with 3 eligible persons, the selected person was a participant. For this household,

$${}_R^1 W_{401} \approx 3055.7035 \,; \qquad W_{hi} = 3 \,; \qquad R\_SP_j \approx 1.03279$$

$${}_R^1 W_i = 3055.7035 * 3 * 1.03279 \approx 9467.67$$

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

5.8.3 Subsample Proportion Adjustment for Person Weight

The Ukraine STEP sample is comprised of three subsample components:
   4) ULMS Panel Sample, i.e., subsample 1, C=1;
   5) New ULMS-2012 subsample i.e., subsample 2, C=2;
   6) STEP Voting Precincts Urban Subsample i.e., subsample 3, C=3.

The participant sample of individuals from each STEP subsample component comprises the overall STEP final sample of participating individuals. Thus, the final sample of individuals from each STEP subsample component is a proportion of the overall STEP final sample of individual participants.

Each component subsample's proportionate share of the overall STEP sample of <u>individuals</u> is hereafter referred to as the Person Subsample Proportionate Adjustment Factor, $^{C}S_{g}$.

The Person Subsample Proportionate Adjustment Factor is calculated <u>at the stratum level</u> as follows:

$$25) \quad {}^{C}S_{g} = \frac{{}^{c}k_{g}}{{}^{1}k_{g} + {}^{2}k_{g} + {}^{3}n_{g}} = \frac{{}^{c}k_{g}}{k_{g}}$$

> where, $^{c}k_{g}$ is the number of participant persons in subsample C in stratum 'g',
> k_{g} is the total number of participant persons in stratum 'g',
> i.e., $k_{g} = {}^{1}k_{g} + {}^{2}k_{g} + {}^{3}k_{g}$

Example 25:   **Person Subsample Proportionate Adjustment Factor**

The Person Subsample Proportionate Adjustment Factors in stratum 1 which has 112 participating households is calculated as follows:

For Subsample 1, in stratum 1, $^{1}S_{1} = \frac{{}^{1}k_{1}}{{}^{1}k_{1} + {}^{2}k_{1} + {}^{3}n_{1}} = \frac{71}{71+17+21} = \frac{71}{109} \approx 0.6514$

For Subsample 2, in stratum 1, $^{2}S_{1} = \frac{{}^{1}k_{1}}{{}^{1}k_{1} + {}^{2}k_{1} + {}^{3}n_{1}} = \frac{17}{71+17+21} = \frac{17}{109} \approx 0.1560$

For Subsample 3, in stratum 1, $^{3}S_{1} = \frac{{}^{1}k_{1}}{{}^{1}k_{1} + {}^{2}k_{1} + {}^{3}n_{1}} = \frac{21}{71+17+21} = \frac{21}{109} \approx 0.1927$

Table 15 provides the Person Subsample Proportionate Adjustment Factor, ^CS_g, by Stratum, i.e., by OBLAST.

Table 15:	Case Summaries: Subsample Number of Participant Persons & Proportion of Participant Persons By OBLAST							
	Subsample *				% of OBLAST Participant Persons (^CS_g)			
OBLAST	1	2	3	Total	1	2	3	Total
1	71	17	21	109	65.1%	15.6%	19.3%	**100.0%**
2	85	24	16	125	68.0%	19.2%	12.8%	**100.0%**
3	14	9	23	46	30.4%	19.6%	50.0%	**100.0%**
4	23	9	29	61	37.7%	14.8%	47.5%	**100.0%**
5	14	6	12	32	43.8%	18.8%	37.5%	**100.0%**
6	109	27	51	187	58.3%	14.4%	27.3%	**100.0%**
7	251	48	99	398	63.1%	12.1%	24.9%	**100.0%**
8	25	6	27	58	43.1%	10.3%	46.6%	**100.0%**
9	5	3	9	17	29.4%	17.6%	52.9%	**100.0%**
10	81	15	20	116	69.8%	12.9%	17.2%	**100.0%**
11	40	11	10	61	65.6%	18.0%	16.4%	**100.0%**
12	42	6	15	63	66.7%	9.5%	23.8%	**100.0%**
13	128	27	30	185	69.2%	14.6%	16.2%	**100.0%**
14	54	12	22	88	61.4%	13.6%	25.0%	**100.0%**
15	34	9	11	54	63.0%	16.7%	20.4%	**100.0%**
16	59	10	39	108	54.6%	9.3%	36.1%	**100.0%**
17	48	12	30	90	53.3%	13.3%	33.3%	**100.0%**
18	30	3	11	44	68.2%	6.8%	25.0%	**100.0%**
19	41	8	30	79	51.9%	10.1%	38.0%	**100.0%**
20	7	6	12	25	28.0%	24.0%	48.0%	**100.0%**
21	133	23	22	178	74.7%	12.9%	12.4%	**100.0%**
22	40	9	15	64	62.5%	14.1%	23.4%	**100.0%**
23	21	5	13	39	53.8%	12.8%	33.3%	**100.0%**
24	41	5	13	59	69.5%	8.5%	22.0%	**100.0%**
25	15		11	26	57.7%	0.0%	42.3%	**100.0%**
26	46	12	19	77	59.7%	15.6%	24.7%	**100.0%**
Total	1457	322	610	2389	61.0%	13.5%	25.5%	**100.0%**

    *   **1-ULMS-2003 Panel**
        2-New ULMS-2012
        3-STEP urban Subsample

5.9 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. It is a comprehensive weight that incorporates the person weight adjustments for all subsamples. This weight is appended to the STEP data file.

The final person weight incorporates the response-adjusted subsample person weight and its proportionate share of the overall STEP sample of individual participants. The final comprehensive person weight is determined by adjusting the final subsample person weight by its Person Subsample Proportionate Adjustment factor, $^{C}S_{g}$,

Therefore, for person 'i' in stratum 'g', the final person weight, $^{C}W_{gi}$, is calculated as follows:

$$26) \quad {}^{C}W_{gi} = {}_{R}^{C}W_{i} * {}^{C}S_{g}$$

> **The final person weight is appended to the STEP data file.**

Example 26: **Final Person Weight**

  i) In stratum 1, selected HH_ID # 614 with 2 eligible persons, the selected person was a participant in subsample 1. For this person, the final person weight is calculated as follows:

  $${}^{1}W_{1i} = {}_{R}^{1}W_{i} * {}^{1}S_{1} = 19613.3732 * 0.6514 \approx 12775.6835$$

  ii) In stratum 1, in participant household # 13736 with 4 eligible persons, the selected person was a participant in subsample 2. For this person,

  $${}^{3}W_{1,13736} = {}_{R}^{3}W_{i} * {}^{3}S_{1} = 110793.315 * 0.1560 \approx 17279.6913$$

  ii) In stratum 1, in participant household # 50114 with 3 eligible persons, the selected person was a participant in subsample 3. For this person,

  $${}^{3}W_{1i} = {}_{R}^{3}W_{i} * {}^{3}S_{1} = 30151.3007 * 0.1927 \approx 5808.9662$$

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 Person Weights to Population Totals

The STEP final person weights were adjusted to the January-2013 national estimates of the urban population by gender and 5 year age groups.

6.1 Benchmark-adjusted Person Weight

6.1.1 Person Weight: Adjustment for Known Gender, Age Group Population

The final person weights based on the combined sample from the three STEP component subsamples are the basis for the STEP estimated number of persons aged 15 to 64 by gender and 5 year age group at the Ukraine national level, $\hat{Y}_{GA}$. The Ukraine Survey Team provided the Excel file 'UKR_stat_2013', which includes corresponding counts, $\ddot{Y}_{GA}$, based on the January-2013 population estimates of the State Committee of Statistics.

In order to make $\hat{Y}_{GA}$ consistent with the corresponding Ukraine January-2013 population estimates, $\ddot{Y}_{GA}$, of the State Committee of Statistics, a final benchmark-adjusted person weight, **W_BM$_i$**, was calculated for each Gender and 5 year Age Group combination GA at the national level:

$$27) \quad W\_BM_i = {}^C W_{gi} * \frac{\ddot{Y}_{GA}}{\hat{Y}_{GA}}$$

Table 16 provides the data that was used to adjust the STEP final person weight to the January-2013 population estimates of the State Committee of Statistics for the STEP target population; it includes, for each Gender and 5 year Age Group combination, the Step estimated urban population and the corresponding January-2013 population estimates of the State Committee of Statistics.

Table 16: STEP Estimated Urban Population vs Ukraine January-2013 population estimates of the State Committee of Statistics by Gender & Age Group

Age Group	# of Sample Cases			STEP Estimated Urban Population Gender_AgeGrp_Est [1]			Ukraine Urban Population 2013 Gender_AgeGrp_2013 [2]		
(5 years)				Male	Female		Male	Female	
	Male	Female	Total	$\hat{Y}_{1A}$	$\hat{Y}_{2A}$	Total	Y_{1A}	Y_{2A}	Total
01 - '15 to 19'	64	73	137	752,804	1,117,078	1,869,882	804,869	774,900	1,579,769
02 - '20 to 24'	87	123	210	1,060,073	1,278,168	2,338,241	1,118,307	1,077,635	2,195,942
03 - '25 to 29'	112	181	293	1,273,989	1,835,598	3,109,587	1,408,147	1,393,752	2,801,899
04 - '30 to 34'	73	140	213	827,460	1,411,012	2,238,472	1,276,839	1,276,951	2,553,790
05 - '35 to 39'	82	134	216	799,224	1,324,047	2,123,271	1,130,939	1,192,554	2,323,493
06 - '40 to 44'	76	120	196	587,121	1,260,776	1,847,897	1,021,499	1,131,628	2,153,127
07 - '45 to 49'	70	148	218	776,212	1,628,927	2,405,139	966,883	1,156,303	2,123,186
08 - '50 to 54'	77	191	268	645,573	1,980,598	2,626,172	1,084,460	1,376,691	2,461,151
09 - '55 to 59'	71	224	295	729,341	2,144,523	2,873,864	931,767	1,277,373	2,209,140
10 - '60 to 64'	92	251	343	719,204	1,639,764	2,358,968	812,407	1,194,214	2,006,621
Total	804	1,585	2,389	8,171,001	15,620,491	23,791,492	10,556,117	11,852,001	22,408,118

1 - 'Gender_AgeGrp_Est' is based on the STEP final Person Weight

2 - 'Gender_AgeGrp_2013' based on January-2013 population estimates of the State Committee of Statistics

7   Weighted Estimates from STEP

7.1   Weighted Household Count – Estimated Number of Target Population Households

The Step estimated number of households in the Ukraine urban area based on the underlined basic household weight before any non-response adjustment is approximately 12,813,654. However, the STEP data collection found that some of the sampled households are ineligible for STEP for reasons such as vacant, not habitable, no eligible household member, etcetera.; therefore, as shown in the third row of Table 17, when ineligible households are excluded, the weighted number of eligible households is **10,304,326**.

Table 17: Ukraine – Weighted Number of Households – STEP Estimates	
Source of Household Count	**Estimated # of Households**
1. STEP, based on the household weight for all sampled households [**all households, i.e., participants, non-participants, ineligibles**]	12,813,654
2. STEP, based on the final household weight for participant households [**eligible households only**]	10,304,326

For participating selected households, Table 18 provides a summary of the central tendency for the final household weight for participating households.

Table 18: Summary - Household Weights	
Statistic	**Final Household Weight**
No. of Participant Households	2453
Mean	4200.70
Std. Error of Mean	63.40
Median	3529.96
Minimum	197.00
Maximum	35679.27
Sum of Weights	10,304,325.84

7.2   Weighted Person Count – Estimated Target Population

The weighted count of the non-institutionalized persons 15 to 64 years of age (inclusive) living in private dwellings in urban areas of **Ukraine** at the time of data collection is provided in Table 19. Based on the 2389 participating selected persons, the national sum of the final person weights from the STEP survey is 23,791,492; the national sum of the benchmark-adjusted final person weights from the STEP survey is 22,408,118.

Table 19:	Ukraine Estimated Number of Urban Persons Aged 15 to 64 by Stratum (i.e., OBLAST)		
Stratum (OBLAST)	No. of Sample Cases	Person Weight: All Subsamples	Person Weight-BM*: All Subsamples
1	109	1,156,135	1,190,734
2	125	2,719,988	2,502,638
3	46	443,615	459,213
4	61	568,026	654,848
5	32	225,559	195,403
6	187	1,912,439	1,879,700
7	398	2,806,423	2,769,559
8	58	647,739	579,093
9	17	362,115	301,056
10	116	1,246,705	1,134,872
11	61	604,545	557,952
12	63	504,848	486,373
13	185	1,463,646	1,220,068
14	88	1,313,774	1,216,218
15	54	612,332	635,799
16	108	1,068,207	934,712
17	90	655,495	568,230
18	44	561,656	582,743
19	79	467,098	416,836
20	25	330,366	258,009
21	178	1,534,944	1,450,888
22	64	559,924	469,198
23	39	382,831	338,844
24	59	642,952	603,660
25	26	399,309	417,113
26	77	600,821	584,358
Total	**2389**	**23,791,492**	**22,408,118**

*   Person Weight, Benchmark (BM) adjusted for National Gender & 5 year Age Group Population

For participating selected persons, Table 20 provides a summary of the central tendency of the number of eligible persons in the participating households.

Table 20: Summary - Person Weights, Eligible Persons			
Statistic	**Eligible Persons in Participant HHs**	**Person Weight: All Subsamples**	**Person Weight-BM*: All Subsamples**
No. of Respondents	2389	2389	2389
Mean	2.2403	9959	9380
Std. Error of Mean	.02097	197	197
Median	2.0000	7286	6644
Minimum Weight	1.00	203	1148
Maximum Weight	8.00	117800	102237
Sum of Weights		**23,791,492**	**22,408,118**

* Person Weight, Benchmark (BM) adjusted for National Gender & 5 year Age Group Population

8   Response Rate

An overall response rate of 60.4% was achieved in the Ukraine STEP Survey. Table 21 provides the detailed percentage distribution by final status code.

Table 21:   Response Rate: Ukraine, Number of Activated Sample Cases by Final Status Code

Case Final Status Code	Subsample *			Total	% of Activated	Response Rate
	1	2	3			
07-Dwelling could not be found, insufficient address info			12	12	0.2%	
08-Vacant dwelling / selected address has no household		40	202	242	4.9%	
09-Dwelling not habitable/destroyed/converted to commercial		8	16	24	0.5%	
12-Panel HH has united with other panel HH	9			9	0.2%	
31-No eligible household member aged 15-64	394	45	219	658	13.4%	
Ineligible HHs *(N.B. % is based on total # of activated HHs)*	**403**	**93**	**449**	**945**	**19.3%**	
01-Refusal-Household	177	143	604	924	18.9%	23.4%
02-Unusual circumstance-Household	11		14	25	0.5%	0.6%
03-No knowledgeable household member found after 3 visits	15		2	17	0.3%	0.4%
04-Temporarily absent/unavailable for field period	42	60	244	346	7.1%	8.8%
05-No competent household member	16	9	20	45	0.9%	1.1%
06-Language problem (i.e., between HH member & Survey Team)			3	3	0.1%	0.1%
11-Panel HH has moved; not located	141			141	2.9%	3.6%
32-Refusal by SP to begin modules 2-7	18	3	1	22	0.4%	0.6%
33-Interview termination-SP unusual circumstance			1	1	0.0%	0.0%
35-Unable to contact SP after 3 attempts			1	1	0.0%	0.0%
36-SP absent for entire survey period	1			1	0.0%	0.0%
40-SP language problem			1	1	0.0%	0.0%
98-Incorrect eligible household member selected	29	3	5	37	0.8%	0.9%
99-Ineligible household member selected	1			1	0.0%	0.0%
Non-respondents (Eligible) *(N.B. % based on total # of Eligible HHs)*	**451**	**218**	**896**	**1565**	**31.9%**	**39.6%**
61-SP unable to read language of General Booklet (GB)	2	1	2	5	0.1%	0.1%
62-SP refusal to begin GB	7		1	8	0.2%	0.2%
63-SP unable to do GB due to unusual circumstance			1	1	0.0%	0.0%
64-SP blindness or visual impairment-unable to do GB	4		2	6	0.1%	0.2%
65-SP physical disability	5	1	1	7	0.1%	0.2%
71-SP looked through GB-unable to read/write language of GB	1			1	0.0%	0.0%
72-SP refused to continue GB after starting	1		3	4	0.1%	0.1%
74-GB terminated - unusual circumstance	2			2	0.0%	0.1%
76-Attempted part or all of GB; <3 correct Core answers	5		1	6	0.1%	0.2%
81-No attempt at any Exercise Booklet (EB) items / refusal to start	22	7	6	35	0.7%	0.9%
82-Interruption too long during EB administration	1			1	0.0%	0.0%
83-Unusual circumstance, EB administration terminated			2	2	0.0%	0.1%
84-Attempt at some EB items, no answers marked	8	3	2	13	0.3%	0.3%
85-Attempt at all EB items, no answers marked	5	2	5	12	0.2%	0.3%
86-Marked some EB items, not all items attempted	19	1	8	28	0.6%	0.7%
87-Marked some EB items, all items attempted	107	19	29	155	3.2%	3.9%
88-Marked all EB items	1268	288	547	2103	42.9%	53.2%
Respondents *(N.B. % is based on total # of Eligible HHs)*	**1457**	**322**	**610**	**2389**	**48.8%**	**60.4%**
Total *Eligible* Households	**1908**	**540**	**1506**	**3954**	**80.7%**	**100.0%**
Total Activated Households	**2311**	**633**	**1955**	**4899**	**100.0%**	
94-HH originally selected in Rural Stratum; data quality issue	56			56		
95-Household not activated			145	145		
Total Sample Cases	**2367**	**633**	**2100**	**5100**		

Subsample *   1-ULMS-2003 Panel
                2-New ULMS-2012
                3-STEP urban Subsample

Appendix I. Description ofthe DataWeighting in UKRAINIAN LONGITUDINAL MONITORING SURVEY 2003 (ULMS 2003).

The weights for extrapolation of the survey results on the population have been calculated and added to data file as additional variable.

1. The stages of sample for weights calculation

Algorithm of the units' selection consisted of the following main stages:

- stratification of the population;
- selection of the territorial units of the first stage;
- selection of the territorial units of the second stage;
- households' selection.

Stratification of the population

Stratification of the population was carried out in order to adequately reflect the main features of administrative-territorial division of Ukraine in the sample, as well as to provide the selection of the units from more homogenous groups by the main characteristics. 51 strata were singled out, which represented all the regions of Ukraine with division by urban and rural area (25 oblasts and Kyiv city). Sample size was distributed by strata proportionally to the population.

Selection of the territorial units of the first stage

On the first stage of sample forming, urban settlements and rural rajons were selected. The units of the first stage were selected with the probability proportional to their size. In total, 97 urban settlements and 54 rural rajons were selected. Total number of selected settlements is 178.
Selection of the first stage was implemented independently in every stratum. Within one oblast, for urban and rural areas corresponding lists of urban settlements and rural rajons were constructed sorted by descending population size. Then there was implemented the selection procedure with probability proportional to the size.

Selection of the territorial units of the second stage

On the second stage, the postal districts were selected in urban and rural areas.

Households' selection

On the third stage, the households were selected. For selecting the households by every selected postal district the street, building, and households' addresses were randomly selected one after another. In accordance with the above-described procedure, 3214 households were selected in 2003. The level of households' participation in the survey was about 66%, on average.

2. Extrapolation of the survey results on the population

For the extrapolation of the survey results on the population, statistical weight was brought in correspondence with every household. The constructed system of statistical weights takes into consideration the sample design, refusal rate for households and separate respondents, available additional information, in particular, data of demographic statistics for age-sex composition of the population, number of households, etc.

Resulting weight for i-household is calculated as product of corresponding weight coefficients by the formula:

$$w_i = w_{Bi} \cdot k_{1i} \cdot k_{2i} \cdot k_{3i}, (1)$$

where w_i – resulting weight of i-household;

w_{Bi} – basic weight of i-household;

k_{1i} – coefficient with regard to the level of households' refusal from participation in the survey;

k_{2i} - coefficient with regard to the level of certain people's refusal from participation in the survey;

k_{3i} – coefficient calculated by the procedure of calibration of the statistical weights.

Basic household's weight is calculated basing on the overall probability of household's selection p_i by the formula:

$$w_{Bi} = 1 / p_i . \qquad (2)$$

For concerned sample design p_i value is defined by the formula:

$$p_i = P_{1i} \cdot P_{2i} \cdot P_{3i}, \qquad (3)$$

where P_{1i} – probability to be selected for urban settlement or rural rajon, where the household is situated;

P_{2i} - probability to be selected for postal district, where the household i is situated;

P_{3i} - probability to be selected for every household within postal district.

Calculation of P_{1i}

For urban settlements in urban area and rural rajons in rural area, the probability of selection P_{1i} by every oblast is calculated by the formula:

$$P_{1i} = a. \frac{M_\alpha}{\sum M_m} , \qquad (4)$$

where a – the number of selected urban settlements and rural rajons;

M_α - population of urban settlement or rural rajon, for which P_{1i} is calculated;

ΣM_m – total population of urban settlements or rural rajons in oblast.

The case is possible when depending on concrete values of a and M_m for certain urban settlements or rural rajons calculated P_{1i} value can be more than 1. Then we take $P_{1i} = 1$, a value is decremented by the number of such cases, and population of such urban settlements or rural rajons is excluded from the total number of corresponding population of the oblast.

Calculation of P_{2i}

The probability of selection of certain postal district P_{2i} is calculated by the formula:

$$P_{2i} = b. \frac{M_\beta}{\sum M_V} , \qquad (5)$$

where b – number of postal districts selected in certain urban settlement or rural rajon, where i-household is situated;

M_β - estimate of population served by the postal district, for which P_{2i} is calculated;

ΣM_V – total population of urban settlement or rural rajon.

In case if the sizes of the postal districts differ insignificantly, formula (5) can be rewritten as follows:

$$P_{2i} = \frac{b}{B} , \qquad (6)$$

where B – total number of postal districts in urban settlement or rural rajon, where i-household is situated.

Calculation of P_{3i}

The probability of selection the households within postal district is calculated by the formula:

$$P_{3i} = \frac{1}{N_U} \cdot \frac{1}{N_B} \cdot \frac{n}{N_{hh}}, \qquad (7)$$

where N_U – the number of the streets served by the postal district, where i-household is situated;

N_H – estimation of the number of dwelling houses on the street;

N_{hh} – estimation of the number of addresses (apartments) in one building;

n – the number of selected households.

It should be noted that product $N_U \cdot N_B \cdot N_{hh}$ is an estimation of the total number of households served by the postal district.

For rural area or private sector in urban area, where one house can be accepted as one address ($N_{hh} = 1$), the formula (7) becomes:

$$P_{3i} = \frac{1}{N_U} \cdot \frac{n}{N_B},$$
(8)

Calculation of the coefficients of taking into account households' refusals from interview.

The coefficient of taking into account households' refusals from interview k_{1i} is calculated by specially constructed cells (classes of weighing). For l - cell this coefficient is defined by the formula:

$$k_{1i} = \frac{\sum\limits_{i=1}^{n} w_{Bi} \cdot \lambda_{li}}{\sum\limits_{i=1}^{n} w_{Bi} \cdot \lambda_{li} \cdot \eta_{li}},$$
(9)

where $\lambda_{li} = \begin{cases} 1, & if\ i \in M_l; \\ 0, & if\ i \notin M_l; \end{cases}$

M_l - a quantity of households of *l*-cell;

$\eta_{ji} = \begin{cases} 1, if\_the\_hosehold\_took\_part\_in\_survey \\ 0, if\_the\_household\_refused\_from\_survey \end{cases}$

$n = \sum\limits_{l} n_l$ - sample size.

Calculation of the coefficients of taking into account respondents' refusals from interview

The coefficients of taking into account respondents' refusals from interview k_{2i} is calculated by every specially constructed *p*-cell by the formula:

$$k_{2i} = \frac{\sum\limits_{i=1}^{n} w_{Bi} \cdot k_{1i} \cdot \lambda_{pi}}{\sum\limits_{i=1}^{n} w_{Bi} \cdot k_{1i} \cdot \lambda_{pi} \cdot \eta_{pi}},$$
(10)

where $\lambda_{pi} = \begin{cases} 1, & if\ p \in M_p; \\ 0, & if\ p \notin M_p;. \end{cases}$

M_p - a quantity of people of p-cell;

$\eta_{ji} = \begin{cases} 1, & if\_the\_person\_took\_part\_in\_survey \\ 0, & if\_the\_person\_refused\_from\_survey \end{cases}$

$$n = \sum_p n_p.$$

Taking into account additional information

To take the available additional information into account, households' weights were calibrated with regard to available official statistical data concerning sex-age composition of population and the number of households, where people of 16-73 years old live, by the regions and types of the settlements.

The procedure of calibration consists in following. On the last stage of calculation of statistical weights they are transformed (corrected) so that the results of the sampling observation give precise estimates of the figures known from other available sources.

From mathematic point of view, the procedure of calibration consists in solving the following equations set:

$$\begin{cases} \sum_{i=1}^{s} \dfrac{(w_{Ci} - w_{Bi} \cdot k_{1i} \cdot k_{2i})^2}{w_{Bi} \cdot k_{1i} \cdot k_{2i}} \to \min; \\ \left(\sum_{i=1}^{s} q_{1i}^{g(1)} w_{Ci} = g_1; ...; \sum_{i=1}^{s} q_{1i}^{g(P)} w_{Ci} = g_P \right)_j^k; \\ \left(\sum_{i=1}^{s} q_{2i}^{m(1)} w_{Ci} = m_1; ...; \sum_{i=1}^{s} q_{2i}^{m(P)} w_{Ci} = m_P \right)_j^k; \\ \left(\sum_{i=1}^{s} w_{Ci} = T \right)_j^k; \\ \left(\sum_{i=1}^{s} a_i w_{Ci} = N \right)_j^k; \end{cases}$$

(11)

where $w_{Ci}, i = 1, ..., s$ - the system of statistical weights after calibration;

$q_{1i}^{g(p)}, i = 1, ..., s$ - the number of women in household i, which belong to the age group p of the region j and type of settlement k, $p = 1, ..., P$; $j = 1, ..., J$; $k = 1, ..., K$;

$q_{2i}^{m(p)}, i = 1, ..., s$ - the number of women in household i, which belong to the age group p;

$a_i, i = 1, ..., s$ - the number of i-household's members;

g_p, m_p, $p = 1, \ldots, P -$ estimates of the number of women and the number of men, which belong to age group *p*;

T – estimate of the number of households of certain region and type of settlement;

N – estimate of the number of population of certain region and type of settlement.

Therefore, additional data in this case are:

- estimates of sizes of sex-age groups by regions and types of settlement;

- estimates of population size by regions and types of settlement;

- estimates of the number of households by regions and types of settlement.

To implement the procedure of calibration, data of demographic statistics for population size, as well as data of All-Ukrainian census by sex-age structure of population were used. The calculation was done by eight economic districts of Ukraine (see Table 1), two settlement types (urban and rural), and also by six sex-age groups (men of 15-34, men of 35-59, men of 60-72, women of 15-34, women of 35-54, women of 55-73).

Table 1 – Composition of economic districts of Ukraine

Economic district	Oblast
Eastern	Poltavska Sumska Kharkivska
Donetsk*	Donetska Luganska
Prydniprovsky	Dnipropetrovska Zaporizka Kirovogradska
Prychornomorsky	Autonomous Republic of Crimea Mykolayivska Odeska Khersonska
Podilsky	Vinnytska Ternopilska Khmelnytska
Central	Kyivska Cherkaska Kyiv city
Carpathian	Zakarpatska Ivano-Frankivska Lvivska Chernivetska
Polisky	Volynska Zhytomyrska Rivnenska Chernigivska