Study Global Questions on Forced Displacement and Jobs

Harmonized Host and Refugee Labor Market Survey in Ethiopia

Sampling Description

Contents

1 Introduction

This document describes the sample of the Harmonized Host and Refugee Labor Market Survey (HHR-LMS) conducted in Ethiopia as part of the study on the impacts of forced displacement on labor markets for host communities. Its main aim is to document the sampling procedures and the procedures for calculating the weight in the survey.

The survey is principally a multi topic survey, with several modules. It comprises a household section, with a roster of all household members. The roster includes questions on basic population characteristics. In addition, there is a separate questionnaire for a randomly selected individual (RSI) aged 18-65 in the household. The RSI questionnaire includes a short experimental component where priming question items are introduced conditional on the characteristics of the respondent in order to explore attitudes and perceptions towards refugees as well as exploring labor market integrations.

Sampling frame

The sample design included 150 initial enumeration areas in Addis Ababa and 79 EAs in Somali region. These EAs were selected using probability proportionate to size where size is measured by the number of households. The enumeration areas were selected based on the sample frame prepared for the population census of Ethiopia planned for 2020 but not implemented due to the COVID pandemic and overall security challenges in the country. The Ethiopian Central Statistical Service (CSS) conducted the selection of the EAs and provided their list along with detailed maps of the areas.

Using maps of the selected enumeration areas provided by CSS, the study team conducted the listing of all households in the selected EAs with door-to-door visits. The listing exercise was carried out during February-March 2022 in Addis Ababa and during May-June 2022 in Somali region by a team of local field workers recruited and trained for this purpose.

In Addis Ababa, we employed adaptive cluster sampling (ACS) to capture enough refugee households. Using the listing of households in the initial 150 clusters in Ababa, we identified those EAs that have 10 percent or more refugee households and conducted the listing of all their neighbors. This resulted in listing additional 71 EAs clusters that are identified as neighbors to these initial clusters. The exercise served as a basis for selection of both refugees and national households in Addis Ababa.

In general, the sample design is a two-stage sample, with an extra third stage for individuals randomly selected in households (RSI). Within each household, one person is selected at random (RSI) from the list of the eligible members: persons aged between 18 and 65 years old in a national household; or refugees aged between 18 and 65 years old in a non-national household.

Sampling refugee populations in urban settings

Refugee populations are generally found in small numbers particularly in urban settings such as that of Addis Ababa. The number of refugees residing in Addis was estimated to be around 72,000 according to the figures from UNHCR[1] and the city had an estimated population of about 5.3 million residents in 2022. These settings can be characterized as constituting "rare populations" where sampling frames are not readily available. To this end, we employed adaptive sampling that refers to designs (Thomson 1997[2]; Thomson and Seber 1996[3]) in which the procedures for selecting units to include in the sample may depend on values of the variable of interest observed during the survey. Such sample designs depend on values of the variable of interest observed during the survey and not before the survey; they therefore pose challenges in the definition of sampling weights, as further detailed below.

Due to the difficulty of surveying refugees in urban settings where they tend to be in smaller proportions compared to the host population, the survey used an adaptive cluster sampling approach to ascertain a large-enough representative sample of refugees would be obtained. The procedure begins by constructing the list of households in the initial randomly selected PSUs. The listing process asked only for the name of the head of household (for identification purposes), the number of household members and nationality. Nationality proved to be an effective proxy for refugee status. A threshold for the number of non-national households in a PSU was set for the listing exercise to be extended to all of the PSU's nearest neighbor PSUs. Theoretically, the listing should then sequentially continue in all neighbors of these additional PSUs, so long as the threshold criteria is fulfilled. In the present survey, a single neighbor listing was conducted for cost reasons but resulted in sufficient number of refugee households. The key element in the effectiveness of adaptive cluster sampling relies on the settlement distribution of refugee households in urban settings. When refugees are settled in clusters with a high-enough concentration, the adaptive cluster sampling is a useful method to efficiently and effectively sample refugee populations.

[1] UNHCR (November 2022). https://data.unhcr.org/en/documents/details/97597
[2] Thompson, S. K. (1997). Adaptive Sampling in Behavioral Surveys. *NIDA Research Monograph , 167, 296-319*.
[3] Thompson, S. K., & Seber, G. A. (1996). *Adaptive Sampling. New York: John Wiley & Sons Inc.*

2 Sampling weight calculations-PPS sample

The sampling design is applied in the localities described above, within the Addis Ababa and Somali regions. The sampling design refers to 229 initial clusters. In general, the sample design is a two-stage sample, with an extra third stage for individuals randomly selected in households (RSI). Within each household, one person is selected at random (RSI) from the list of the eligible members: persons aged between 18 and 65 years old in a national household; or refugees aged between 18 and 65 years old in a non-national household.

2.1 Notation

To describe the sample precisely and calculate inclusion probabilities, we need to introduce some notation. This is done in Table 1. In general, the notation uses subscripts to indicate the sample stage, and superscripts to indicate the source of the data used. Thus $N_{h,c}$ means the number of units in stratum h, cluster c.

Table 1: Notation used

Symbol	Meaning	
N	Household count as recorded in the 2014 Census	
N^{l}	Household count as listed from fieldwork	
n	Number of households selected in a PSU	
m	Number of PSUs selected in a stratum	
P	Inclusion probability	
W	Weight	
l	Index of sampling information collected from listing	superscript
h	Index of stratum	subscript
c	Index of primary sampling unit (PSU)	subscript
i	Index of household	subscript
rsi	Index of an individual selected in a household	subscript
d	Index of design weight	superscript
s	Index of sampling weight	superscript
r	Index of relative weight	superscript

2.2 Inclusion probability

Inclusion probability of PSUs

The enumeration areas are the primary sampling units (PSUs). The inclusion probability for a PSU *c* in stratum *h* is the following (for all strata except the initial enumeration area stratum):

Equation 1 Inclusion probability for PSU

$$p_{h,c} = \frac{N_{h,c} * m_h}{N_h}$$

Inclusion probability of households

Within each PSU, all households are listed, and a fixed number of households are then selected to be surveyed. The inclusion probability for a household *i* within a PSU *c* in stratum *h* is then the following:

Equation 2 Inclusion probability for households

$$p_{h,c,i} = \frac{n_{h,c}}{N^{l}_{h,c}}$$

The overall inclusion probability for a household then becomes:

Equation 3 Overall inclusion probability for household

$$p_i = p_{h,c} * \ p_{h,c,i}$$

Inclusion probability of RSI

The sampling units are classified into two categories: national and non-national households. A household with a household head who declared during the listing that they are a national of the host country are henceforth referred as a 'national household'. A household with a household head who declared that they are not a national of the host country are henceforth referred as a 'non-national household'. Within each household, one person is selected randomly (RSI) from the eligible members: persons aged between 18 and 65 in a national household; or refugees aged between 18 and 65 in a non-national household). The inclusion probability for an RSI within a household is then:

Equation 4 Inclusion probability for individual

$$P_{rsi} = \frac{1}{N^{rsi}}$$

The final inclusion probability for an RSI is then the product of the household inclusion probability and the RSI inclusion probability.

Equation 5 Overall inclusion probability for individual

$$P_{i,rsi} = p_i * \ p_{rsi}$$

2.3 Derivation of weights

We calculate both design weights and sampling weights (or survey weights) for both households and individuals in this project. The design weight of a sampling unit (household or individual) is the inverse of the overall probability with which the unit was selected in the sample. The sampling weight of a sampling unit is the design weight corrected for non-response or other calibrations. The design weight is the basic weight for the survey. Other weights are calculated based on the design weight.

Design weights

Thus, the design weights for a household or an individual are:

Equation 6 Design weight for a household

$$W_i^d = \frac{1}{p_i}$$

Equation 7 Design weight for an individual in a household

$$W_{i,rsi}^d = \frac{1}{p_{i,rsi}}$$

We usually do not use the design weights directly in estimating survey results. It is because the design weights are adjusted for non-response, as discussed below.

Sampling weights

Non-response and non-response corrections

Generally, one may distinguish between two types of non-response: unit non-response and item non-response. Unit non-response pertains to the non-response of a whole unit, such as a household. In that case, almost nothing is known about that household. Item non-response relates to the lack of information on a specific item for a unit; for instance, a person does not answer questions about income. Here we will only consider the unit non-response.

Unit non-response and non-response correction

The results of the interviews or attempted interviews can be studied using a classification frame of non-response, derived by Hidiroglou, Drew, and Gray (Hidiroglou, 1993). The response categories in the framework are given below.

The framework is built around the observation that an interview can be missing for two reasons. First, it may be that the selected household or individual does not exist or that it does not belong to the sampling frame because it does not meet the eligibility criteria. Second, a selected household or individual, which exists and is eligible, may refuse, or not be found at home. Also, the classification has to

take into account that there will be some situations where the interviewer cannot determine if a household exists or not.

Interviewers also sometimes encounter the situation that the household or individual is available for an interview but cannot provide any useful information because the respondent is sick or otherwise incapable of answering. The cases with interview status not determined may still have the possibility of being interviewed. Therefore, they are reallocated proportionally between the category non-response and no interview possible.

The categories used to identify the results of visits of this project are not precisely the same as the framework. The response categories are then adjusted as in Table 2 and Table 3.

Table 2 Response category – household questionnaire

Code	Category	Household
1	Interview completed	Response
2	Refusal converted	Response
3	Partially completed	Non-Response
4	Refusal	Non-response
5	Dwelling unit vacant	No interview possible
6	No Eligible person	Non-response
7	Dwelling doesn't exist in the area	No interview possible
8	No contact	Non-response
9	Status not determined	Not clear, distributed between Non-response and No Interview possible

Table 3 Response category – RSI questionnaire

Code	Category	RSI
1	Interview completed	Response
2	Refusal converted	Response
3	Partly completed	Non-response
4	Refusal	Non-response
5	Not at home	Non-response
6	Incapacitated	Non-response
7	Person not eligible	Not clear, distributed between Non-response and No Interview possible

Non-response always occurs in survey efforts. Nevertheless, since the extent and seriousness of non-response vary in the field, the plan of non-response correction must be reconsidered after the fieldwork. The rates of non-response correction can be computed from the above framework, as indicated below.

Weight adjustment and sampling weight

When there is unit non-response, direct use of the design weight will result in biased estimation. The biases generally take two forms. One is that when totals are estimated with the design weight, the total will be too small because non-response implies that units that should have been added to the total are missing. The other form is that estimation of statistics other than the total may be biased because non-responding units may have particular characteristics.

One way to reduce the biases produced by unit non-response is to adjust the design weights. The method of correcting the weights for non-response is the so-called "adjustment cell method" (Lehtonen & Pahkinen, 2004; Little & Rubin, 2002). In this approach, households or individuals that are considered to be fairly similar are identified and grouped. The non-response rate is calculated for each group, called adjustment cells.

The inverse of the non-response rate in each adjustment cell is then used to adjust the design weight for each household or individual. The result is the so-called sampling weight. The weighted sample size is now as it would have been if all households or individuals had responded. Using the sampling weights also increases the relative contribution to the estimates of units that are similar to those missing.

In the HHR-LMS project, the adjustment cells used are strata (the three regions). The non-response correction is carried out separately for the household and the RSI responses.

The correction factor to the weights for non-response is given below with notations in Table 4.

Table 4: Notation for non-response adjustment

Symbol	Meaning	
C	Adjustment (Correction) factor	
U	Number of sampling units	
a	Index of adjustment cell	subscript
c	Index of cluster adjustment	subscript
i	Index of household adjustment	subscript
rsi	Index of individual adjustment	subscript
r	Index of responding household/individual	superscript
f	Index of non-responding household/individual	superscript
l	Household count from listing	superscript
m	Household count as recorded in the 2014 Census but missing from the listing	superscript

The number of possible interviews (i.e., the denominator in the non-response rate) is the sum of categories 1, 2, 3, 4, 6, and 8 in Table 2 for household interviews; the sum of categories 1, 2, 3, 4, 5, and 6 in Table 3 for individual interviews. The number of non-respondent units is the sum of categories 3, 4, 6, and 8 in Table 2 for household interviews; categories 3, 4, 5, and 6 in Table 3 for individual interviews. The Status Not Determined category is proportionally allocated to the non-respondent category.

Equation 8 Non-response correction ratio

$$C_{a,*} = \frac{1}{\frac{U^r_{a,*}}{U^r_{a,*} + U^f_{a,*}}}$$

Twelve clusters were selected but not listed due to practical issues with regards to accessing the localities within the listing period. Adjustment is made based on the number of households recorded in the 2014 census.

Equation 9 Non-response correction ratio for missing clusters

$$C_{a,c} = \frac{1}{\frac{U^l_a}{U^l_a + U^m_a}}$$

The weights are then adjusted according to the following equations:

Equation 10 Weight correction according to non-response

$$W^s_i = W^d_i * C_{a,i} * C_{a,c}$$

$$W^s_{i,rsi} = W^s_i * \frac{1}{p_{rsi}} * C_{a,rsi}$$

Extreme weights correction

Truncation is also done with extreme weights. We decided that the max of weights should not be higher than the sum of the median weight and two standard deviations of weights. The Truncation of the weights is done within regions. The distribution of weights is skewed right, with some extremely large weights but no extremely small weights. Small weights are therefore not truncated.

Relative weights

Two types of weights can be used for weighing household or individual data. The sampling weights reflect the size of the population of interest. While the relative weights (normalized sampling weights) retain the sample size and only adjust the relative contribution of each unit of analysis (PSU, household, or individual). The normalized standard weight of a sampling unit is calculated based on its sampling weight by multiplying the sampling weight with a unique constant. The constant is the total number

of completed cases divided by the total number of weighted cases (based on the sampling weight). With the normalized sampling weight, the number of unweighted cases coincides with the number of weighted cases for both total households and total individuals.

To prevent large numbers for the number of weighted cases in data analysis and confusion about the assumed population size, we normalize the sampling weights by region. We calculated the relative weights for both households and RSI respondents with interview status "complete interview" or "Refusal converted".

Equation 11 Relative weight

$$W_i^r = W_i^s * \frac{n_i}{\sum W_i^s}$$

$$W_{rsi}^r = W_{rsi}^s * \frac{n_{rsi}}{\sum W_{rsi}^s}$$

3 Sampling weight calculations- ACS sample

In addition to the PPS sampling method described above, data was also collected using an alternative sampling design, namely adaptive cluster sampling (ACS) in Addis Ababa, Ethiopia. The ACS sampling design was based on 150 initial clusters and 71 network clusters. It is a two-stage sample, with an extra third stage for individuals randomly selected in households (RSI).

3.1 Notation

To describe the sample precisely and calculate inclusion probabilities, we need to introduce additional notation building on Table 1 shown above.

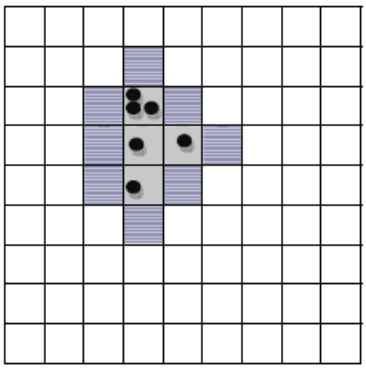
Table 5: Notation used

Sym-bol	Meaning	
N	Household count as recorded in the 2014 Census	
N^l	Household count as listed from fieldwork	
n	Number of households selected in a PSU	
m	Number of PSUs selected in a stratum	
P	Inclusion probability	
W	Weight	
l	Index of sampling information collected from listing	superscript
h	Index of stratum	subscript
c	Index of primary sampling unit (PSU)	subscript
k	Index of network	subscript
i	Index of household	subscript
rsi	Index of an individual selected in a household	subscript
d	Index of design weight	superscript
s	Index of sampling weight	superscript
r	Index of relative weight	superscript

The definition of a network is essential for calculating the inclusion probabilities for PSUs in the ACS approach. A network is defined to include an initial PSU and the PSUs that have been selected together with the initial PSU as a result of the application of the adaptive cluster sampling process. This is illustrated in Figure 1, which shows a grid of PSUs where some have the target population, e.g., refugee households. If one of the PSUs with the target population was sampled initially (e.g., the one with three dots, corresponding to three households), the evenly shaded units correspond to its neighbors; while the cross-hatched units correspond to the neighbors with no refugee households present. The

neighbors with no members of the target population present are called edge units. The network for probability estimation is the network resulting from a selection of an initial PSU, but with edge units removed.

Figure 1 Network illustration



3.2 Inclusion probability

Inclusion probability of PSUs

When adaptive cluster sampling is applied, it is not strictly an inclusion probability, but an intersection probability of networks that is used to calculate weights. The intersection probability, which is equal for networks and units, is expressed as follows:

Equation 12 Intersection probability for networks

$$P_{h,c} = 1 - \left[1 - \frac{N_{h,k}}{N_h}\right]^m$$

$N_{h,k,}$ and N_h now refer to the number of households in the network and the stratum, respectively. Note that the above formula is strictly speaking only valid when cluster inclusion is stopped at stratum boundaries (or where there is no stratification). Selection for each stratum is then independent, and strata can be treated in the usual way (Thompson & Seber, 1996).

Inclusion probability of households

Within each PSU, a fixed number of households are to be selected. The inclusion probability for a household within a PSU c in stratum h is then the following:

Equation 13 Inclusion probability for households in a cluster

$$p_{c,i} = \frac{n_{h,c}}{N_{h,c}^{l}}$$

Note the use of the superscript *l*, which indicates that it is the listed number of households that should be used rather than the initial estimate of households from the census. And the refugee households and the national households are listed separately. The overall inclusion probability for a household then becomes:

Equation 14 Overall inclusion probability for households

$$P_i = P_{h,c} * P_{c,i}$$

Inclusion probability of RSI

Within each household, one person is selected randomly from the eligible members: persons aged between 18 and 65 in a national household; or refugees aged between 18 and 65 in a non-national household). The inclusion probability for an RSI within a household is then

Equation 15 Inclusion probability for RSI respondent

$$P_{rsi} = \frac{1}{N_{rsi}}$$

The final inclusion probability for an RSI is then the product of the household inclusion probability and the RSI inclusion probability.

Equation 16 Overall inclusion probability for individual

$$P_{i,rsi} = p_i * \ p_{rsi}$$

3.3 Derivation of weights

As before, we calculate both design weights and sampling weights (or survey weights) for both households and individuals in this project. The design weight of a sampling unit (household or individual) is the inverse of the overall probability with which the unit was selected in the sample. The sampling weight of a sampling unit is the design weight corrected for non-response or other calibrations. The design weight is the basic weight for the survey. Other weights are calculated based on the design weight.

Design weights

Thus, the design weights for a household or an individual are:

Equation 16 Design weight for a household

$$W_i^d = \frac{1}{p_i}$$

Equation 17 Design weight for an individual in a household

$$W_{i,rsi}^{d} = \frac{1}{p_{i,rsi}}$$

We usually do not use the design weights directly in estimating survey results. It is because the design weights are adjusted for non-response, as discussed below.

Sampling weights

Non-response and non-response corrections

Generally, one may distinguish between two types of non-response: unit non-response and item non-response. Unit non-response pertains to the non-response of a whole unit, such as a household. In that case, almost nothing is known about that household. Item non-response relates to the lack of information on a specific item for a unit; for instance, a person does not answer questions about income. Here we will only consider the unit non-response.

Unit non-response and non-response correction

The results of the interviews or attempted interviews can be studied using a classification frame of non-response, derived by Hidiroglou, Drew, and Gray (Hidiroglou, 1993). The response categories in the framework are given below.

The framework is built around the observation that an interview can be missing for two reasons. First, it may be that the selected household or individual does not exist or does not belong to the sampling frame. This is the case when a household or individual is not considered eligible for a survey. Second, a selected household or individual, which exists and is eligible, may refuse or not be found at home. The classification must also consider that there will be some situations where the interviewer cannot determine whether a household exists.

Interviewers also sometimes encounter the situation that the household or individual is available for an interview but cannot provide any useful information because the respondent is sick or otherwise incapable of answering. The cases with interview status not determined may still have the possibility of being interviewed. Therefore, they are reallocated proportionally between the category non-response and no interview possible.

The categories used to identify the results of visits of this project are not precisely the same as the framework. The response categories are then adjusted as in Table 2 and Table 3

Table 6 Response category – household questionnaire

Code	Category	Household
1	Interview completed	Response
2	Refusal converted	Response
3	Partially completed	Non-Response
4	Refusal	Non-response
5	Dwelling unit vacant	No interview possible
6	No Eligible person	Non-response
7	Dwelling doesn't exist in the area	No interview possible
8	No contact	Non-response
9	Status not determined	Not clear, distributed between Non-response and No Interview possible

Table 7 Response category – RSI questionnaire

Code	Category	RSI
1	Interview completed	Response
2	Refusal converted	Response
3	Partly completed	Non-response
4	Refusal	Non-response
5	Not at home	Non-response
6	Incapacitated	Non-response
7	Person not eligible	Not clear, distributed between Non-response and No Interview possible

Non-response always occurs. Nevertheless, since the extent and seriousness of non-response vary in the field, the plan of non-response correction must be reconsidered after the fieldwork. The rates of non-response correction can be computed from the above framework, as indicated below.

Weight adjustment and sampling weight

When there is unit non-response, direct use of the design weight will result in biased estimation. The biases generally take two forms. One is that when totals are estimated with the design weight, the total will be too small because non-response implies that units that should have been added to the total are missing. The other form is that estimation may be biased because non-responding units may have particular characteristics.

One way to reduce the biases produced by unit non-response is to adjust the design weights. The method of correcting the weights for non-response is the so-called "adjustment cell method" (Lehtonen & Pahkinen, 2004; Little & Rubin, 2002). In this approach, households or individuals that are considered to be fairly similar are identified and grouped. The non-response rate is calculated for each group, called adjustment cells.

The inverse of the non-response rate in each adjustment cell is then used to adjust the design weight for each household or individual. The result is the so-called sampling weight. The weighted sample size is now, as it would have been if all households or individuals had responded. Using the sampling weights also increases the relative contribution to the estimates of units that are similar to the missing ones.

In the HHR-LMS project, the adjustment cells used for household weights are strata combined with the type of household (national or non-national). The adjustment cells used for RSI weights are strata. The non-response correction is carried out separately for the household and the RSI responses.

The correction factor to the weights for non-response is given below with notations in Table 4.

Table 8: Notation for non-response adjustment

Symbol	Meaning	
C	Adjustment (Correction) factor	
U	Number of sampling units	
a	Index of adjustment cell	subscript
c	Index of cluster adjustment	subscript
i	Index of household adjustment	subscript
rsi	Index of individual adjustment	subscript
r	Index of responding household/individual	superscript
f	Index of non-responding household/individual	superscript
l	Household count from listing	superscript
m	Household count as recorded in the 2014 Census but missing from the listing	superscript

The number of possible interviews (i.e., the denominator in the non-response rate) is the sum of categories 1, 2, 3, 4, 6, and 8 in Table 2 for household interviews; the sum of categories 1, 2, 3, 4, 5, and 6 in Table 3 for individual interviews. The number of non-respondent units is the sum of categories 3, 4, 6, and 8 in Table 2 for household interviews; categories 3, 4, 5, and 6 in

Table 3 for individual interviews. The Status Not Determined category is proportionally allocated to the non-respondent category.

Equation 18 Non-response correction ratio

$$C_{a,*} = \frac{1}{\frac{U^r_{a,*}}{U^r_{a,*} + U^f_{a,*}}}$$

Nine clusters were selected but not listed due to practical challenges during listing. No adjustment is made for that.

The weights are then adjusted according to the following equations:

Equation 19 Weight correction according to non-response

$$W_i^s = W_i^d * C_{a,i}$$

$$W_{i,rsi}^s = W_i^s * \frac{1}{p_{rsi}} * C_{a,rsi}$$

Extreme weights correction

Truncation is also done with extreme weights. We decided that the max of weights should not be higher than the sum of the median weight and two standard deviations of weights. The Truncation of the weights is done within regions. The distribution of weights is skewed right, with some extremely large weights but no extremely small weights. Small weights are therefore not truncated.

Relative weights

Two types of weights can be used for weighing household or individual data. Sampling weights reflect the size of the population of interest, while relative weights (normalized sampling weights) maintain the sample size and only adjust the relative contribution of each unit of analysis (PSU, household, or individual). The normalized standard weight of a sampling unit is calculated based on its sampling weight by multiplying the sampling weight with a unique constant. This constant is the total number of completed cases divided by the total number of weighted cases (based on the sampling weight). With the normalized sampling weight, the number of unweighted cases aligns with the number of weighted cases for both total households and total individuals.

To prevent large numbers for the number of weighted cases in data analysis and confusion about the assumed population size, we normalize the sampling weights by region. We calculated the relative weights for both households and RSI respondents with interview status "complete interview" or "Refusal converted".

Equation 20 Relative weight

$$W_i^r = W_i^s * \frac{n_i}{\sum W_i^s}$$

$$W_{rsi}^r = W_{rsi}^s * \frac{n_{rsi}}{\sum W_{rsi}^s}$$

4 References

Hidiroglou, M. A., J. D. Drew and G. B. Gray (1993). A framework for measuring and reducing nonresponse in surveys.

Lehtonen, R. and E. Pahkinen (2004). Practical methods for design and analysis of complex surveys. Chichester, West Sussex, England ; Hoboken, NJ, J. Wiley.

Little, R. J. A. and D. B. Rubin (2002). Statistical analysis with missing data. Hoboken, N.J., Wiley.

Thompson, S. K., & Seber, G. A. (1996). *Adaptive Sampling. New York: John Wiley & Sons Inc.*