Using Addresses as Sampling Units in the 2007 Health Information National Trends Survey

Greg Norman[1] and Richard Sigman[1]
[1]Westat, 1600 Research Blvd., Rockville MD 20850

Abstract
The Health Information National Trends Survey (HINTS) is a recurring household survey that collects data about the sources and characteristics of health information[2] from a representative sample of all adults in the United States. Data for the 2007 HINTS were collected by telephone from an RDD sample of phone numbers and by mail from a sample of addresses. This paper focuses on the address sample, which was selected from a sampling frame based on the U.S. Postal Service's Delivery Sequence File and supplemented with non-USPS sources to increase coverage in rural areas. In this paper we discuss the address sample stratification, the household and person-level weights, rates of non-delivery and response by type of address, the handling of multiple ways a household can receive mail, and the characteristics of the supplemental list of addresses.

Key Words: HINTS, Delivery Sequence File, mail survey

1. Introduction

The National Cancer Institute and its contractor, Westat, conducted the Health Information Trend Survey (HINTS) in 2003, 2005, and 2007. HINTS data is used to determine the sources and characteristics of health information of adults living in U.S. households. In 2003 and 2005, random-digit dialing (RDD) was used to collect the HINTS data. In 2007, a dual-frame design was used by adding an address based sample frame to the RDD frame.

This paper focuses on the address based sampling frame used in HINTS 2007. Section 2 discusses the sample design of the address sample. Section 3 discusses weighting of the address sample respondents. Section 4 provides results and properties of the address sample. Section 5 discusses the potential biases of not augmenting the address sample with a supplement list of addresses. Section 6 concludes with lessons learned that can be applied to future cycles of HINTS.

2. Design of the HINTS 2007 Address Sample

In this section we discuss the design of the HINTS 2007 address sample. Section 2.1 describes the mail survey design and the associated mailing protocols. Next, in section 2.2, we discuss the selection of the sampling frame and its properties. This is followed by section 2.3 which describes a supplemental sample of addresses that were used. In section

---

[2] Topics covered by HINTS include access to and usage of health information, the levels of trust that HINTS respondents have in different types of health information, and how HINTS respondents would like to receive health information in the future.

2.4 we talk about how the sample was stratified. In sections 2.5 and 2.6, respectively, we describe the household and within household reporting units.

2.1. Mail Survey Design

Households sampled in HINTS 2007 were sent an advance letter introducing the study and explaining the questionnaires that they would receive. This letter included a set of FAQs on the back. A week later, households were sent a package of three questionnaires with a request that each adult in the household return a completed questionnaire. This package included a $2 incentive. Two weeks after the initial mailing, households that had not yet responded were sent a postcard reminding them to return the completed questionnaires. Two weeks after the reminder postcard, households that had not yet responded were sent a second package of questionnaires by FedEx. For post office box addresses, this second mailing was sent using USPS priority mail.

2.2. Sampling Frame

The address sample for HINTS 2007 was obtained from Marketing Systems Group (MSG). The decision to use MSG was based largely on the result of an evaluation study conducted by Link et al. (2005). In this study they compared five address vendors in terms of the coverage of their lists for a six-state area. Three vendors had high levels of under-coverage in one or more states. Of the remaining two vendors, only MSG could provide sampling services for a single-stage sample of addresses while the other vendor required two stages of sampling - first the sampling of carrier routes and then the sampling of individual addresses. Compared to a single-stage design, the two-stage design was less desirable as it would have been more costly and less precise.

The MSG frame of addresses is updated bimonthly from the USPS's Delivery Sequence File. The DSF is an electronic data product that USPS provides to address vendors to update their address lists (USPS, 2006). MSG receives DSF updates to its address lists for ZIP Codes in which the number of addresses in MSG's lists is at least 90 percent but not more than 110 percent of the number of addresses in the ZIP code according to USPS.

The DSF, with the exception of general delivery, contains the most current information on mailing addresses serviced by the USPS. This information is available for the following types of addresses:

- Addresses that currently receive or have received mail delivery.

- Addresses on city routes to which carriers do not deliver because of alternative delivery arrangements, e.g. to post office boxes. These addresses are referred to as "throwbacks".

- Addresses on city routes vacant longer than 90 days and likely to be long-term vacancies, which are not considered seasonal. These addresses are referred to as "vacants".

- Addresses delivered seasonally (no DSF information is available, however, on the dates of the mailing season). These addresses are referred to as "seasonals".

Link et al. (2005) evaluated the coverage of the MSG address frame for six states. For each of the counties in this six-state study area, they compared the number of addresses on the MSG frame to Census Bureau estimates and they defined high undercoverage

counties as those where the MSG frame was 10 percent or more deficient, when compared to Census estimates. Their results showed that in counties with less than a 25 percent urban population, nearly 90 percent had high undercoverage. While for counties with 75 percent or more of the population living in an urban area, less than 5 percent had high undercoverage.

The sampling frame for HINTS 2007 was all non-vacant residential addresses in the United States. The sampling unit was an individual address. The MSG frame allowed for the matching of names and telephone phone numbers for addresses, but because the sampling unit was an address, the names and telephone number of individuals believed to reside at the sampled address were not used to contact the household.

2.3. Augmenting the Sampling Frame

One of the problems in using an address-based sampling frame is the presence of "simplified" addresses. This type of addresses is an address that does not contain a street address or box number. Therefore, unless the name of someone in the household is used there is insufficient detail for mailing. For HINTS 2007, alternative address sources were used to augment the sampling frame to account for these simplified addresses.

MSG was able to obtain additional information for the simplified addresses through secondary sources. The weighted proportion of all addresses in the final sample that were identified through this supplemental procedure was 2.5%. Although this procedure only provided a small percentage of the overall weighted sample size, the percentage in rural areas was much higher. In non-metropolitan counties (as defined by Beale codes[3] 4 through 9) we found that 9.0 percent of addresses were from the augmented list; whereas in the metropolitan counties (Beale codes 1 through 3) the proportion of addresses that came from the augmented list was only 1.1 percent.

2.4. Stratification

The sampling frame was stratified into two strata - a high-minority stratum and a low-minority stratum. Strata were formed by first using demographic data from Claritas to classify Census block groups based on their population percentages of Hispanics and African Americans. Block groups were then matched to addresses by their 9 digit zip code. A report was then produced indicating the number of households at varying levels of Hispanic and African American concentration. Addresses in block groups that had a population proportion for Hispanics or African Americans that equaled or exceeded 24 percent were assigned to the high-minority stratum. All other addresses were assigned to the low-minority stratum. A profile of the two sampling strata is shown in Table 1.

Table 1: Profile of the sampling strata in HINTS 2007

Stratum	Proportion of frame	Coverage of African Americans and Hispanics	Prevalence of African American or Hispanic in Stratum
High minority	25.1%	71.9%	62.7%
Low minority	74.9%	28.1%	8.2%

---

[3] http://www.ers.usda.gov/Briefing/Rurality/RuralUrbCon/

An equal-probability sample of addresses was selected from each stratum. The date of the sample selection was July 12th, 2007. In order to obtain a large enough sample for subgroup analysis, the high minority stratum was oversampled so that the proportion of the sample coming from the high minority stratum was 50 percent.

2.5. Household Reporting Units
The frame chosen for the HINTS 2007 address sample contained duplicate units because some households have multiple ways in which they receive mail. To permit adjustment for duplication in the sampling frame, a question was asked on the mail questionnaire to measure the number of ways the responding household receives mail. This is discussed in greater detail in section 4.4.

2.6. Within Household Reporting Units
Within a household all adults that live at a given address were asked to complete a questionnaire. Hence, the mail sample was a stratified cluster sample, in which the household was the cluster. Our decision to not subsample the adults in selected households was the result of an evaluation study conducted by Battaglia et al. (2005, 2008). Their study compared three respondent-selection methods for household mail surveys: (1) any adult in the household; (2) the adult in the household having the next birthday; and (3) all adults in the household. The results from this study showed that the next birthday and all-adults methods yielded household-level completion rates that were comparable to the any-adult method - the method that the researchers assumed to have the least respondent burden. In addition, these results showed that differences in response rates by gender and age were less for the all-adults methods than for the other two methods.

3. Weighting

Each responding adult received a weight consisting of five major components; base weights, household nonresponse adjustment, number of ways the household receive mail, within household nonresponse adjustment, and a calibration adjustment. Each of these components is described below.

3.1. Base Weights
The base weight is the reciprocal of an addresses probability of selection, which depends on the stratum the address was selected from. The cases in the high-minority stratum had a base weight of approximately one-third the base weight of the addresses in the low-minority stratum - reflecting the oversampling of high-minority areas.

3.2. Household Nonresponse Adjustment
The base weights of the households that returned at least one questionnaire were adjusted to reflect nonresponse by the remaining households (minus those addresses returned as having bad addresses). Eight nonresponse cells were created based on cross-classifications of Census Region and Stratum (high vs. low minority). The overall adjustment factor for this adjustment was 2.50.

3.3. Ways Households Receive Mail
There may be more than one way that a household receives residential mail. Adjustments were made based on the number of ways that a household reported receiving mail. This

adjustment is analogous to the multiple telephone adjustment for an RDD survey. The mean adjustment for the number of ways a household receives mail was 0.95.

3.4. Within Household Nonresponse Adjustment

Each household was asked to have all adults in the household fill out a questionnaire and return it. However, for numerous households the number of questionnaires returned did not match the number of adults reported. Therefore, the weights were adjusted to reflect this within-household nonresponse. The mean adjustment factor for within-household nonresponse was 1.29.

3.5. Calibration Adjustments

The purpose of calibration is to reduce the sampling variance of estimators through the use of reliable auxiliary information (see, for example, Deville & Sarndal, 1992). In the ideal case, this auxiliary information takes the form of known population totals for particular characteristics (called *control totals*). Even if the auxiliary information contains sampling error, however, calibration will reduce the sampling variance of resulting estimators as long as the sampling errors of the auxiliary information are significantly smaller than those of the survey itself[4].

The American Community Survey (ACS), conducted by the U.S. Census Bureau, has much larger sample sizes than those of HINTS. The ACS estimates of population totals have lower sampling error than the corresponding HINTS estimates, making calibration of the survey weights to ACS control totals beneficial.

Calibration variables were selected among those that were on the ACS public-use file. The variables used were correlated to important HINTS outcomes. The following ACS characteristics correlated well with key HINTS questionnaire items:

- Age
- Gender
- Educational Attainment
- Marital Status
- Race
- Ethnicity
- Census Region

One of the recommendations of research conducted by Han and Cantor (2007) into nonresponse bias in the 2005 HINTS was that in addition to characteristics from the ACS, health-related variables should also be used for calibration. Two variables from the 2006 National Health Information Survey (NHIS) that correspond to questions asked in the HINTS survey were used. They were:

- Percent With Health Insurance
- Percent Ever Had Cancer.

Raking to the control totals for these variables (either alone or cross-classified with each other) was then performed. As a result of raking HINTS weights to the control totals,

---

[4] Calibration was done using full-sample auxiliary totals for control variables. Consequently the amount of variance reduction is slightly overstated because we did not reflect the variability of the control totals.

estimates calculated from HINTS data for the control-total variables agree with those calculated from the source data for the control totals. For example, the national-level estimate of Percent Ever Had Cancer calculated from HINTS 2007 data agrees with the estimate calculated from NHIS 2006 data.

4. Results

Several key results are discussed in this section. First, in section 4.1 we evaluate the composition of the address frame used in HINTS 2007. In sections 4.2 and 4.3, respectively, we look at the non-delivery and response rates. In section 4.4 we look at our question which measured the number of ways households can receive mail. Finally in section 4.5 we evaluate the impact of our stratification scheme.

4.1. Frame Composition

In addition to the information needed for mailing (e.g. street name and number, city, state, and zip code), each address contained additional classification variables. Among these was the type of delivery point. There are four types of delivery points: city delivery, rural, post office (PO) box, and highway contract.

Although not used in the stratification, the type of delivery point for an address can provide an indication of the HINTS 2007 sample composition. This can then be compared to 2007 USPS totals (http://www.usps.com/financials/_pdf/annual_report-2007.pdf).

Using the HINTS 2007 base weight and comparing our results to the 2007 USPS totals, we find that the HINTS 2007 sample slightly over represent city and rural delivery points, while under representing PO boxes and highway contract delivery points. These results are shown in Table 2.

Table 2: Delivery Points in HINTS 2007 Compared to 2007 USPS Totals

	City delivery	*Rural*	*PO Box*	*Highway contract*
HINTS 2007	60.6%	29.3%	8.5%	1.6%
USPS 2007	59.0%	27.5%	11.6%	1.8%

4.2. Household Level Non-delivery Rates for HINTS 2007

Overall the non-delivery rates for the HINTS 2007 address sample were relatively low. The weighted non-delivery rate was 8.8%. For the low minority stratum the non-delivery rate was lower than for the high minority stratum. While the non-delivery rate for the supplemental sample was greater than for the main sample. The greatest difference in non-delivery rates was by type of delivery point, ranging from a low of 7.6% for rural deliveries to a high of 21.9% for highway contracts. These results are shown in Table 3.

Table 3: HINTS 2007 Non-delivery Rates

	Non-delivery rate
Overall	8.8%
High minority stratum	10.6%
Low minority stratum	8.2%
Main sample	8.6%
Supplemental sample	16.6%
City delivery	8.3%
Rural delivery	7.7%
PO boxes	13.9%
Highway contract	21.9%

4.3. Response rates for HINTS 2007

Looking at the household-level response rate, the weighted response rate for HINTS 2007 address sample was 40.0%. In the low minority stratum the response rate was higher than in the high minority stratum. Also, the supplemental sample had a higher response rate than for the main sample. Finally in looking at the rates by type of delivery point, the rates ranged from a low of 28.6% for post offices boxes, to a high of 44.9% for rural routes. A possible explanation of the lower response rate for P.O. boxes is that for non-P.O. box addresses FedEx was the delivery method for the second mailing, whereas for PO boxes, USPS priority mail was used for the second mailing.

The final within household response rate was 77.7%. This number is fairly consistent across groups. However for three groups their rates were less than seventy-five percent. These were, the supplemental sample (73.1%), highway contract (71.4%), and (69.8%).

Overall person-level response rates were computed as the product of the household and within-household rates. The overall person-level response rate was 31.1%. The high minority stratum had a lower person-level response rate than the low minority stratum. Looking at the type of sample (main vs. supplemental) the response rates are very similar, with the supplemental sample having a slightly higher rate. Finally, comparing the type of delivery points, PO boxes and highway contracts had fairly low rates when compared to city or rural delivery. The difference in response rate across delivery points suggests that it would be a useful variable to use when creating non-response cells.

The response rates for the various groups in the HINTS 2007 address sample are shown in Table 4.

Table 4: Response Rates for the HINTS 2007

	Household	*Individual*	*Overall*
All respondents	40.0%	77.7%	31.1%
High minority stratum	28.4%	78.1%	22.2%
Low minority stratum	43.8%	77.5%	33.9%
Non-augmented sample	39.9%	77.8%	31.1%
Augmented sample	44.2%	73.1%	32.3%
City delivery	39.4%	78.8%	31.0%
Rural delivery	44.9%	77.4%	34.7%
PO boxes	28.6%	69.8%	20.0%
Highway contract	31.7%	71.4%	22.6%

4.4. Analysis of the ways households receive mail

As mentioned in Section 2.5, some household may receive mail multiple ways. To measure this, the following question was included in the mail questionnaire:

At which of the following types of addresses does your household currently receive residential mail? Mark all that apply

A street address with a house or building number
An address with a rural route number
A U.S. Post office box (P.O. Box)
A commercial mail box establishment (such as Mailboxes are Us, Mailboxes etc.)

The weighted mean number of ways that households receive mail by group is shown in Table 5. The number of ways for receiving mail is slightly larger for the high minority stratum (1.06 vs. 1.03, p = 0.002). There was no significant difference between the main and supplemental sample. Across delivery point, the number of ways that households receive mail was not significantly different for three types of delivery: city delivery, street delivery, and highway contracts. However, post office boxes had a significantly higher mean, 1.24. This result suggest that it may be desirable to revise this question placing an increased focus on determining whether the household has a post office box that it uses for receiving residential mail.

Table 5: Number of Ways for Receiving Mail

	Mean # of ways for receiving mail
Overall	1.04
High minority stratum	1.06
Low minority stratum	1.03
Non-augmented sample	1.04
Augmented sample	1.03
City delivery	1.02
Rural delivery	1.04
PO boxes	1.24
Highway contract	1.02

4.5. Stratification

For the HINTS 2007 address sample, the sample was stratified into high and low minority strata. In doing this there were competing goals. One, we wanted to provide a large enough pool of interviews for the African Americans and Hispanics domains and secondly we wanted to reduce the variance of the estimates for these groups. In this section we evaluate the effectiveness of our stratification in meeting these goals.

First we will look at how effective our stratification scheme was in identifying African American or Hispanic households. These results are shown in Table 6. The high minority strata accounts for 71.0% of all African Americans or Hispanics, even though it contains just 25.1% of all household. Within the high minority stratum, 58.7% of respondents identified themselves as an African American and/or a Hispanic (31.3% African American and 29.7% Hispanic). For the low minority stratum, only 10.1 percent identified themselves as an African-Americans or Hispanic (4.3% African American and 6.2% Hispanic). These results are very similar to the results in Table 1, which were calculated from Claritas data at the Census block group. This indicates that our

stratification scheme was affective in identifying African American and Hispanic households.

Table 6: Results of Stratification

Domain	High minority	Low minority
Overall percent of African Americans or Hispanics	71.0%	29.0%
Strata percent African American or Hispanic	58.7%	10.1%
African American	31.3%	4.3%
Hispanic	29.7%	6.2%

Next we looked at the impact of our stratification on the variances of resulting estimates. Table 7 looks at variance reduction (or increase) due to oversampling the highest minority stratum. The variance reduction factor was calculated for 11 questions in the HINTS 2007 data, using the following formula:

$$\text{variance reduction factor} = \frac{\text{Var}_{\text{Stratified}}}{\text{Var}_{\text{Proportional}}},$$

where $\text{Var}_{\text{Stratified}}$ is the stratum variance using our stratified design and $\text{Var}_{\text{Proportional}}$ is the estimated stratum variance had we used proportional allocation. (The Appendix provides additional details on how we calculated the variance reduction factor.) A factor of less than 1.00 indicates a reduction in variance while a factor greater than 1.00 would indicate an increase in variance due to stratification. The first row of Table 7 shows that for the entire population, oversampling the minority stratum results in a minimal increase in the overall sampling variance; whereas, for the two minority groups of interest, there is a substantial reduction.

Table 7: Variance Reduction due to Stratification

Domain	Average Reduction	Minimum Reduction	Maximum Reduction
All households	1.08	0.92	1.24
African American	0.68	0.57	0.76
Hispanic	0.72	0.62	0.91

5. Potential Bias of Not Using the Supplemental Sample of Addresses

In this section we take a look at response to some survey questions to determine whether there are any significant differences between the main and supplemental samples. As discussed in section 2, only 2.5% of address came from supplemental sources. However these supplemental sources contained a disproportionate number of rural addresses. In addition, as discussed in section 4, this supplemental sample had higher non-delivery rates. However, the overall response rates for the main and supplemental sample were not significantly different. But do the supplemental address cases differ in their responses to key survey questions?

To answer this question we analyzed a subset of questions from the HINTS 2007 questionnaire. In total 47 variables were evaluated and 8 variables showed significant differences. These are shown in Table 8. Based on these results there is an indication that

the supplemental cases are different enough from the main sample that to exclude them from the sample would lead to a non-coverage bias.

Table 8: Differences between the Main and Supplemental Samples in the HINTS 2007 Address Sample

	Main	Supplement	p
Means			
Confidence that you could get health related advice if needed	2.72	2.47	0.031
Self-reported health rating	2.47	2.36	0.029
Proportions			
Heard of the American Cancer Society	86.51	95.41	0.009
Participated in any physical activity during the past month	67.45	43.55	0.011
Smoked at least 100 cigarettes in lifetime	47.09	73.38	0.001
Felt worthless in the past 30 days	28.12	45.36	0.033
College graduate or more	24.96	13.81	0.009
Income 50K or more	49.17	20.42	<0.001

6. Summary and Conclusions

In this paper we have looked at the results of using an address sample for HINTS 2007. This sample was provided by MSG and it contained a list of households from the USPS DSF file. This list was then supplemented to include households with simplified addresses, which were primarily found in rural areas. Our final frame of addresses was then stratified into two groups, based on the concentration of African American or Hispanic households.

The HINTS 2007 sample yielded high delivery rates and compared favorably to USPS estimates based on type of delivery point. However it did slightly over represent city and rural delivery while under representing post office boxes. When looking at response rates, overall post office boxes had the lowest response rates. This result may have been confounded by the fact that in the second mailing all other groups were sent their questionnaire via FedEx, while PO boxes had their second-mailing questionnaires sent by USPS priority mail. The difference in response rates across the different types of delivery points suggests that the type of delivery point should be used in creating non-response adjustment cells.

In looking at the number of ways that households receive mail, most households receive mail one way. The notable exception is for post office box households, who are more likely to receive mail a second way. Together these results may indicate that, while the post office boxes where underrepresented in the sample and had the lowest response rates, we are potentially reaching these households through the other types of delivery points.

We also looked at the effect of stratification. The results of this showed that stratifying addresses by percent African American or Hispanic was not only a good way to identify these households but it reduced the variance of estimates for these groups without causing a large increase in the variance for all respondents. In addition, we looked at the

effect of including the supplemental sample of addresses to see whether their exclusion could result in bias. Our analysis showed that significant differences exist between respondents in the main and supplemental samples and that to exclude these cases could lead to non-coverage bias.

References

Battaglia, M.P., Link, M.W., Frankel, M.R., and Mokhad, A.H. 2005. An evaluation of respondent selection methods for household mail surveys. In *JSM Proceedings,* Survey Research Methods Section. Alexandria, VA: American Statistical Association. 2727-2731.

Battaglia, M.P., Link, M.W., Frankel, M.R., Osborn, L., and Mokhad, A.H. 2008. An evaluation of respondent selection methods for household mail surveys. *Public Opinion Quarterly,* 72(3), 459-463.

Deville, J.C., and Sarndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association,* 87, 376-382.

Han, D., and Cantor, D. (2007), A Comparison of Level of Effort and Benchmarking Approaches for nonresponse Bias Analysis of an RDD Survey. In *JSM Proceedings,* Survey Research Methods Section. Alexandria, VA: American Statistical Association. 3155-3160.

Link, M.W., Battaglia, M.P., Giambo, P., Frankel, M.R., Mokhad, A.H., and Rao, S.R. 2005. Assessment of address frame replacements for RDD sampling frames. Paper presented at the 2005 American Association for Public Opinion Research Annual Conference, Miami, FL. Available at
www.m-sg.com/Reference_Library/Address_Frame_Replacements.htm

United States Postal Service, 2006. CDS User Guide. Retrieved April 20, 2006, from
http://ribbs.usps.gov/files/CDS/CDSug.pdf

Appendix

The results in the first row of Table 7 are for all households, whereas the results in the second and third rows are for the analysis domains of all African American households and all Hispanic households, respectively. For the results in the first row of Table 7, we used the following formula to calculate the variance reduction factor for the estimated proportion of adults in all households that belong to a particular HINTS item category:

$$\text{variance reduction factor} = \frac{\sum_{h=1}^{2} W_h^2 P_h \left(1 - P_h\right)\left(\text{deff}_h\right) / r_h}{\sum_{h=1}^{2} W_h^2 P_h \left(1 - P_h\right)\left(\text{deff}_h\right) / r_h^{(prop)}} ,$$

where

p_h = estimated proportion of the adults in all households belonging to the item category,

deff_h = the estimated design effect for the item category in stratum h,

r_h = the number of respondents that are assigned to stratum h,

$r_h^{(prop)}$ = the estimated number of respondents that would be assigned to stratum h if the sample had been proportionately allocated, and

W_h = the proportion of the address frame assigned to stratum h.

For the domain-level results in rows two and three of table 7, we used the following formula:

$$\text{variance reduction factor}_d = \frac{\sum_{h=1}^{2} W_{dh}^2 P_{dh}\left(1-P_{dh}\right)\left(\text{deff}_{dh}\right)/r_{dh}}{\sum_{h=1}^{2} W_{dh}^2 P_{dh}\left(1-P_{dh}\right)\left(\text{deff}_{dh}\right)/r_{dh}^{(prop)}} ,$$

where

p_{dh} = estimated proportion of the adults in the domain belonging to the item category,

$\textit{deff}_{dh}$ = the estimated domain-level design effect for the item category in stratum h,

r_{dh} = the number of respondents in the domain that are assigned to stratum h,

$r_{dh}^{(prop)}$ = the estimated number of respondents in the domain that would be assigned to stratum h if the sample had been proportionately allocated, and

$W_{dh} = \sum_{i \in d} w_{hi} / \sum_{h=1}^{2} \sum_{i \in d} w_{hi}$ = estimated proportion of the domain assigned to stratum h,

where

w_{hi} = the adjusted sampling weight for respondent i in stratum h.