Revisiting Nested Stratification of Primary Sampling Units

Tom Krenzke and Wen-Chau Haung
Westat, 1600 Research Boulevard, Rockville, MD, 20850

Stratified multi-stage cluster area sample designs are used widely when conducting large, in-person surveys in the US because they are cost effective and efficient. In such surveys, it is expensive to conduct a listing operation to create a frame of dwelling units and to travel interviewers to the selected households. Therefore, in the first stage of selection, a usual approach is to form, stratify, and select Primary Sampling Units (PSUs), for example, counties or groups of counties. These geographic areas are formed to reduce interviewer travel costs and to increase the heterogeneity within PSUs.

Prior to selection, PSUs are stratified into homogeneous groups in order to reduce the anticipated sampling variation in the resulting survey estimates. Another objective is to form strata close-to-equal in population totals to help achieve close-to-equal interviewer workloads and a self-weighting sample. Reducing the variation among stratum-level population totals for one-PSU per stratum designs helps to reduce the variance in survey estimates, especially totals, and also reduces the bias in variance estimates. While some stratification designs are conducted to serve multiple purposes or surveys, we focus the discussion on a single survey with a single variable of interest. Stratification searches have been implemented using sophisticated multivariate clustering algorithms, such as described in Friedman and Rubin (1967), Jewitt and Judkins (1988) and a more computer-intensive approach as presented in Ludington (1992). Kish (1965) discusses much effort while implementing a stratification approach, and questions the benefits of such expensive efforts. He mentions that stratification attempts that appear to be very different often lead to about the same variances.

One purpose of this paper is to investigate Kish's conclusive remarks. We describe searches under a simplified multivariate algorithm using nested stratification, likely used by Kish, which attempts to increase homogeneity (using distance measures) and reduce the variation among substrata population totals, while arriving at explicit boundaries, as some may prefer for documentation and clearly communicating the stratification results. In effect, we revisit the efforts undertaken by Kish and his colleagues, measuring the variation across hundreds of stratification schemes.

A second purpose of this paper is to present an evaluation of the PSU stratification design for the 2003 National Assessment of Adult Literacy (NAAL), which used a nested stratification procedure. Subsequent to the conduct of the survey, extensive modeling led to the identification of key variables that would be good stratification variables (such as Decennial Census data) for the future. Also, model-based estimates of low-literacy at the county level have been produced and are used here as an evaluation variable for computing the between PSU variance for different substratification schemes. Lastly, improvements to Westat's PSU stratification software (WesStrat) have allowed more stratification schemes to be created.

Key steps leading up to the substratification process

The underlying scenario for this discussion is to select a stratified probability proportionate to size sample of PSUs that will lead to a self-weighting design. We consider the following steps in the stratification of PSUs.

Determining the measure of size. The measure of size used to select PSUs is typically the population count within the PSUs. The measure of size is used to allocate the total number of substrata proportionate to size to each major stratum. It is also used in forming the substrata, for instance, to reach the objective of equal-sized strata.

Identifying self-representing (SR) PSUs. Self-representing (SR) PSUs are typically PSUs with the largest values of the measure of size. The SR PSUs come into the sample with probability equal to one. Each SR PSU is in a stratum by itself and therefore is excluded from the substratification process.

Determining the number of PSUs and strata. The number of PSUs to select depends primarily on cost and reliability considerations, which includes the increase to sampling variance due to clustering individuals within sampling units. In general, the more PSUs selected, the less clustering but the higher cost due to interviewer travel. Once the number of PSUs is established, the total number of strata is derived by the number of PSUs planned in the sample and the number of sample PSUs per stratum. Once the self-representing PSUs are identified, under a one-PSU per stratum design, the total number of strata is equal to the total number of PSUs needed, which is equal to the number of SR PSUs and non-self representing (NSR) PSUs. Therefore, the number of NSR strata is equal to the total number of strata minus the number of SR PSUs. Under a two-PSU per stratum design, the total number of strata is equal to the number of SR PSUs added to one-half the number of NSR PSUs needed. Therefore, the number of NSR strata is equal to the total number of strata minus the number of SR PSUs.

Identifying major strata. The non-self representing (NSR) PSUs on the frame are grouped into major strata. The major strata are typically formed to ensure representation across geographic areas while allowing for estimates to be reported for the domains they represent (e.g., state-level estimates). Once identified, they serve as hard boundaries when forming the substrata. The major strata should also be related to the survey outcome measure.

Identifying substratification variables. Typically 2 to 4 variables are used to form strata within the major strata. The substratification variables should be related to the survey outcome variable of interest. They may be selected after processing a stepwise regression, or after review of literature of past analyses. Some examples of stratifiers used in demographic surveys include median household income, total population size, and proportion of population with a college degree.

Allocating the total number of NSR strata to the major strata. The total number of NSR strata needs to be allocated to the major strata. When the allocation is done proportionate to the measure of size, strata totals may be more equal in size across all strata. With a one-PSU per stratum design, it is preferable to have an even number of strata allocated to each major stratum, since strata will need to be combined (paired) to facilitate variance estimation.

Substratifying each major stratum. Once the steps described above are completed, the substratification process is implemented. The next section discusses a nested stratification approach that has been used in several surveys at Westat, including the Early Childhood Longitudinal Study, NAAL, and the Adult Literacy and Lifeskills Survey, each sponsored by the National Center for Education Statistics, and the National Center for Health Statistics' National Health and Nutrition Examination Survey**.**

A nested stratification approach

Given the dual objective of reducing the between PSU variance and arriving at close-to-equal measure of size totals across substrata, we undertook a search for the best substratification solution given the underlying nested stratification approach. We should note that we treat the objectives as equal in this paper, however, in practice, one may be favored over the other depending on the situation. The nested stratification design arrived at begins with forming substrata from one stratifier. With a 2nd stratifier, substrata are formed within each stratum from the 1st stratifier. A 3rd stratifier is used to form substrata within each substrata formed by the 2nd stratifier, and so on. The splitting on each stratifier is focused on arriving at close-to-equal size totals across substrata. This nested approach can be thought of in terms of a tree structure, where a set of branches is created by splitting the set of PSUs into groups. The branches are identified by using weighted percentiles on the measure of size (MOSVAR). The percentiles are weighted by measure of size. For example, suppose percent black (*PCT_BLK)* is the lone stratifier ($SV = 1$) to form three substrata $H_g = 3$. Then there is only one possible solution for the tree structure. Given that solution, two ($H_g - 1$) cutpoints are created on the stratifier *PCT_BLK*. To find the cutpoints, it first sorts the PSUs by *PCT_BLK,* and then computes the cumulated sum of measure of size for each subsequent PSU record. The cutoffs in *PCT_BLK* are the points that contribute 1/3 and 2/3 of the total measure of size. Appendix A provides more details of the algorithm. Given the number of allocated substrata to the major stratum g (H_g), and given the number of stratifiers (SV), all possible nested substratification schemes are found under the above splitting approach. The number of possible substratification schemes (Z) is equal to $Z = SV^{H_g - 1}$. Table 1 shows the number of schemes related to the number of stratifiers and number of substrata.

Table 1. Number of substratification schemes by number of stratifiers (*SV*) and number of substrata

Number of Substrata	Number of schemes where *SV*=1	Number of Schemes where *SV*=2	Number of Schemes where *SV*=3	Number of schemes where *SV*=4
2	1	2	3	4
3	1	4	9	16
4	1	8	27	64
5	1	16	81	256
6	1	32	243	1024
7	1	64	729	4096
8	1	128	2187	16384
9	1	256	6561	65536
10	1	512	19683	262144
11	1	1024	59049	1048576
12	1	2048	177147	4194304
13	1	4096	531441	16777216
14	1	8192	1594323	67108864

The underlying approach is illustrated in Figure 1, which shows the nested stratification charts for three substrata (H=3) and three stratifiers (*SV*=3), referred to as *a*, *b*, and *c*. The Figure shows scheme notations, for example, at the top left chart (1): (1,1,1)(1,1,2)(1,1,3) – giving 3 nodes, one for each final substratum. This scheme provides very useful information; for example, the first position in each node is related to stratifier 1, the second position is related to stratifier 2, and the third position is for stratifier 3. For chart (1), the scheme notation is shown as (1,1,1)(1,1,2)(1,1,3), because the first two positions are constant, and the third position changes, reflecting that only the third stratifier is used in the stratification. For chart (2), the scheme notation (1,1,1)(1,1,2)(1,2,1) shows that the first stratifier is not used, and the second stratifier is split into two, with one further split made on the third stratifier.

After each automated substratum solution is generated, the evaluation tools (between PSU variance and equal-size strata measure) are computed. The objective is to reduce the values of these measures when grouping PSUs into strata. The computation for the between-PSU variance for an evaluation variable's total *U*, for major stratum *g* and substratum *h* among the total number of PSUs *I* is as follows:

$$BETWVAR_{gh} = \sum_{i=1}^{I_h} PROB_{ghi} \times (\hat{U}_{gh(i)} - U_{gh})^2$$

where, $PROB_{ghi} = \dfrac{MOSVAR_{ghi}}{MOSVAR_{gh}}$

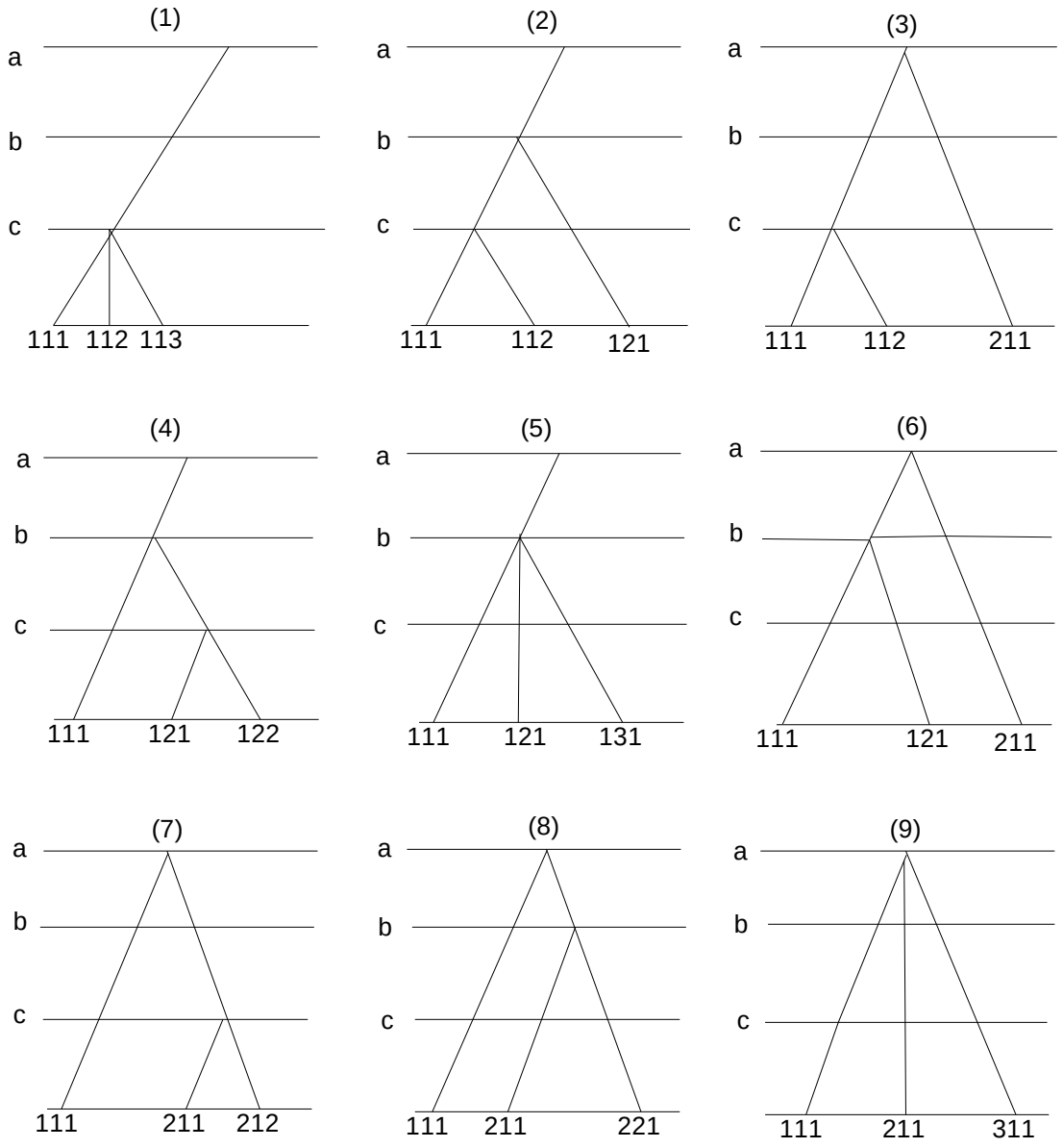
where, $MOSVAR_{ghi}$ = size measure for PSU *i*, substratum *h* for major stratum *g*,
$MOSVAR_{gh}$ = size measure for substratum *h* for major stratum *g*,

where, $\hat{U}_{gh(i)} = \dfrac{U_{ghi}}{PROB_{ghi}}$ = estimated total of evaluation variable U_{gh} by PSU *i* of substratum *h*, major stratum *g*,

where U_{ghi} = total of evaluation variable for PSU *i*, substratum *h* for major stratum *g*,
U_{gh} = total of evaluation variable for substratum *h* for major stratum *g*.

For each major stratum *g*, the between PSU variance becomes: $BETWVAR_g = \sum_{h=1}^{H_g} BETWVAR_{gh}$. If the evaluation variables are in terms of a percentage, they are converted to totals *U*.

Figure 1. Nested Stratification Charts for Three Strata (H=3) and Three Stratifiers (SV=3)



The equal-size strata measure is simply the variance of the substratum-level MOSVAR values. It is computed as follows for major stratum *g*:

$$ESS_g = \frac{\sum_{h=1}^{H_g} \left(MOSVAR_{gh} - \frac{\left(\sum_{h=1}^{H_g} MOSVAR_{gh} \right)}{H_g} \right)^2}{H_g - 1}$$

Evaluating the 2003 NAAL substratification scheme

We use the substratification process outlined above to evaluate the 2003 NAAL stratification scheme. As described in Mohadjer et al (2009), the NAAL 2003 household study was designed to be a nationally representative sample from the 50 states and the District of Columbia of persons in households or college dormitories who were 16 years of age or older at the time of interview. The NAAL sample was selected based on a four-stage area sample design, aimed at reducing the cost of interviewing and assessing respondents in their homes. The first stage of selection was of primary sampling units (PSUs). PSUs were defined to be counties or sets of counties with the following general characteristics: 1) PSUs were required to have a minimum population of 15,000 persons; 2) PSUs were required to be no wider than 100 miles in maximum point-to-point distance; 3) PSUs consisted of counties that were either all Metropolitan Statistical Area (MSA) or non-MSA; and 4) PSUs were required to stay within state boundaries.

A total of 1,884 PSUs were formed and combined into 100 strata. A total of 100 PSUs was selected (one-per stratum) with probability proportionate to size as the first-stage sample, with the estimated size equal to the year 2000 population. Associated with the NAAL design were six state-level samples, called the State Assessment of Adult Literacy (SAAL). An additional 74 PSUs were sampled for the SAAL states of which 14 overlapped with the 84 national NSR PSUs. To simplify the evaluation, we focused only on the stratification relating to the national NAAL sample.

The 16 PSUs with largest measures of size (based on the total household population from the 2000 Decennial Census) were identified as self-representing. Twelve of these 16 PSUs were identified as having probabilities equal to one and the remaining four PSUs had initial probabilities of selection close to 1 and were also selected as self-representing. Each of the SR PSUs was treated as a single stratum and the remaining PSUs were stratified into 84 NSR strata.

The stratification process for the NSR PSUs started with the formation of 17 major strata defined by Census Division and MSA status, where non-MSA PSUs in Census Divisions[1] 1 and 2 were combined into one major stratum. Then, the sample size of 84 NSR PSUs was allocated proportional to the total measure of size in each of the major strata. Table 2 presents the allocation of 84 NSR PSUs among the 17 major strata. As it is desirable for the purpose of variance estimation to select an even number of PSUs from a major stratum, the allocated numbers were mostly rounded to even numbers.

Table 3 presents the variables used for substratification within each major stratum. The variables used in the substratification process were identified earlier by performing a regression analysis with the demographic variable relating to the percentage of the population that were high school graduates 25 years and older. The variables were listed in the table in order of importance, relating to the variability explained by each variable (as measured by R-square) in the regression analysis. The 2003 NAAL substratification process was done using the nested approach described above.

To evaluate the NAAL PSU stratification scheme, the ideal evaluation variable would be one that is an outcome variable from the survey itself, and is available for each county in the entire country. After the 2003 NAAL, county-level estimates were produced using small area estimation (SAE) techniques that rely on NAAL survey data, as well as data from other sources, such as the Decennial Census. As described in Mohadjer et al (2009), NCES undertook the project to produce estimates of adults at the lowest literacy level for individual counties using statistical modeling approaches (such as in Rao 2003). The local area predictions estimate the percent lacking basic prose literacy skills (*BPLS*). These model-dependent estimates are called "indirect" estimates to distinguish them from standard or "direct" estimates that do not depend on the validity of a statistical model. The SAE approach uses the NAAL direct estimates and the modeling, to borrow strength from other counties and uses the auxiliary data to help improve upon the imprecision of available direct estimates. We use the indirect estimates to evaluate the stratification scheme in terms of the between PSU variance.

[1] The nine census divisions are: 1) New England, 2) Middle Atlantic, 3) East North Central, 4) West North Central, 5) South Atlantic, 6) East South Central, 7) West South Central, 8) Mountain, 9) Pacific.

Table 2. Allocation of NSR PSUs in major strata

Census Division	MSA status	No. of PSUs	Population	Allocation	
				Exact	Rounded
1	MSA	13	8,569,586	3.31	3
1+2	Non-MSA	88	5,262,752	2.03	2
2	MSA	36	19,721,290	7.61	7
3	Non-MSA	238	8,874,076	3.42	3
3	MSA	63	28,319,740	11.49	12
4	Non-MSA	272	7,376,807	2.85	3
4	MSA	36	11,274,000	4.35	4
5	Non-MSA	264	10,215,210	3.94	4
5	MSA	73	35,349,252	13.63	14
6	Non-MSA	211	6,853,688	2.64	3
6	MSA	31	9,696,238	3.74	4
7	Non-MSA	220	6,649,947	2.57	3
7	MSA	50	16,668,660	6.43	6
8	Non-MSA	132	4,457,347	1.72	2
8	MSA	29	10,293,970	3.97	4
9	Non-MSA	72	3,604,643	1.39	2
9	MSA	40	23,113,905	8.92	8
Total	NSR	1868	216,301,111	84	84
Total	SR	16	55,868,955	16	16
Total	All PSUs	1884	273,643,259	100	100

Note: Sums may not add to totals because of rounding.

Table 3. Substratification variables used in NAAL PSU stratification

Division	**MSA Status**	**Substratification Variables**
1, 2, 8 and 9	Non-MSA	Per capita income
3 and 4	Non-MSA	Per capita income, Percent Non-Hispanic White
5, 6, and 7	Non-MSA	Per capita income, Percent Non-Hispanic Black
1 and 2	MSA	Per capita income, Percent Hispanics
3 and 6	MSA	Per capita income, Percent Non-Hispanic Black
4	MSA	Per capita income
5 and 7	MSA	Per capita income, Percent Non-Hispanic Black, Percent Hispanic
8 and 9	MSA	Per capita income, Percent Non-Hispanic White

To gain maximum benefit from stratification, a high correlation between the stratifiers and the key survey outcome variable is required. The correlation between the indirect estimates with the SAE key predictors, as well as the substratification variables used in the 2003 NAAL, are shown in Table 4. The correlation coefficients for the 2003 NAAL stratification process variables are slightly lower as a group when compared to the 2003 NAAL SAE predictors. Also provided in Table 4 are R^2 values for logistic regression models. As we would expect, the R^2 value for the model with the NAAL stratifiers is much lower (0.669) than for the model with the NAAL SAE predictors (0.898). The resulting R^2 value for a model that excludes the percentage of the population below the 150 percent poverty line from the set of NAAL SAE predictors is 0.868 -- still much larger than the model than includes the NAAL stratifiers. The resulting R^2 value for a model that excludes two variables (the percentage who are Black or Hispanic, and the percentage of the population below the 150 percent poverty line) from the set of NAAL SAE predictors is 0.634, which is about the same level as the model than includes the NAAL stratifiers.

Table 4. Logistic regression R^2 values and correlation coefficients with percent lacking *BPLS* and the 2003 NAAL stratifiers and SAE predictors

Covariate	R^2	Correlation coefficients with percent lacking *BPLS*
2003 NAAL stratifiers	0.669	
Per capita income		-0.35
Percentage of the population who are Non-Hispanic White		-0.73
Percentage of the population who are Non-Hispanic Black		0.51
Percentage of the population who are Hispanic		0.56
2003 NAAL SAE predictors – Evaluation stratifiers	0.898 (0.868[a]) (0.634[b])	
Percentage of the population who are foreign-born stayed in the United States 0-20 years		0.45
Percentage of persons age 25 and older with a high school education or less		0.51
Percentage of the population who are Black or Hispanic		0.80
Percentage of the population below the 150 percent poverty line		0.66

[a] This is the resulting R^2 value for a model that excludes the following predictor: Percentage of the population below the 150 percent poverty line.
[b] This is the resulting R^2 value for a model that excludes the following predictors: Percentage of the population who are Black or Hispanic, percentage of the population below the 150 percent poverty line.

The variables used as key predictors in the SAE models are used as stratifiers in the evaluation, excluding the percentage of the population below the 150 percent poverty line. Due to the computer intensive search, for major strata with more than 10 substrata ($H_g > 10$), we used two stratifiers ($SV = 2$). For major strata with $H_g \leq 8$, we used $SV = 4$. As shown in Table 2, no major strata had $H_g = 9$ or 10. Furthermore, the automated approach provides all possible schemes, and sometimes schemes include substrata with just one PSU. For this comparison, we exclude any substratification scheme with at least one substratum with just one PSU.

Since the evaluation stratifiers were predictors used in the SAE model that generated the evaluation measure (indirect estimates), the evaluation is a tough test for the 2003 NAAL scheme. This is exemplified by the results in Table 5, which shows the percentiles on the between PSU variance distribution within each major stratum for the NAAL 2003 stratification scheme among all generated evaluation schemes. The percentiles for non-MSAs, based on between PSU variance, ranged from the 53rd percentile in census division 3 to the 100th (worst scheme compared to the 16 evaluation schemes) in census division 7, and for MSAs, the percentiles ranged from the 8th percentile in census division 1 to the 98th percentile in census division 5. We would expect better results for the equal sized strata measure since it does not depend on the evaluation variable, which is associated with the evaluation stratifiers. For non-MSAs, the NAAL percentiles based on the equal size strata measure ranged from the 12th percentile in census division 7 to the 80th percentile in the combined census divisions 1 and 2, where as for MSAs the percentiles ranged from 0.2 in census divisions 3 and 9, to 46th in census division 1. Each scheme was ranked within the major stratum, both in terms of the between PSU variance measure, and the equal size strata measure. Using the two ranks, the average combined rank among the two measures was then computed and percentile within

each major stratum associated with the NAAL scheme among the average combined ranks associated with the evaluation schemes are also shown in Table 5. For non-MSAs, the 2003 NAAL stratification's percentile ranged from 40 to 80, while for MSAs, the percentiles ranged from 0.3 to 60.

Table 5. Percentiles for 2003 NAAL results by evaluation measure and major strata

Census Division	MSA Status	Number of substratification evaluation schemes	Percentiles		
			Between PSU variance	Equal size strata	Average combined rank
1	MSA	12	7.7	46.2	23.1
1+2	Non-MSA	4	60.0	80.0	80.0
2	MSA	4015	91.7	1.1	45.5
3	Non-MSA	16	52.9	29.4	41.2
3	MSA	1680	46.0	0.2	17.1
4	Non-MSA	16	58.8	70.6	70.6
4	MSA	64	93.8	21.5	60.0
5	Non-MSA	64	93.8	13.8	47.7
5	MSA	723	98.5	3.5	0.3
6	Non-MSA	16	64.7	35.3	52.9
6	MSA	64	40.0	23.1	15.4
7	Non-MSA	16	100.0	11.8	52.9
7	MSA	1024	38.6	9.3	13.2
8	Non-MSA	4	80.0	40.0	40.0
8	MSA	64	89.2	3.1	40.0
9	Non-MSA	4	80.0	40.0	60.0
9	MSA	16269	90.4	0.2	43.4

Improvements to stratification for future adult literacy surveys

The above analysis shows that there can be improvements made to PSU stratification in future adult literacy surveys. The variables used in the extensive search for predictor variables in the SAE model are leading candidates for stratifiers. Using a measure such as the average combined rank, described above, will help to find the best scheme in reducing the between PSU variance and measure of size variation across substrata. We also investigated if it was beneficial to use more stratifiers. The 2003 NAAL scheme included one to three stratifiers for any given major stratum. Table 6 shows the percent relative differences between the optimal solutions among the evaluation measures (separately for between PSU variance and equal size strata measure) for *SV*=2 and for *SV*=4. For MSAs in divisions 3 and 5, only schemes with *SV*=2 were generated and therefore the comparison between two and four stratifiers was not made. The two stratifiers used are the best two predictors in the SAE model and the schemes having four stratifiers are based on the four SAE predictors. A decrease signifies a reduction in the measure when going from two stratifiers to four stratifiers. When comparing the minimums (lowest resulting measure), there are seven major strata with more than a 10% reduction in the between PSU variance (ranging from 0% to -23%), while most strata have more than a 10% reduction in the equal size strata measure. These results imply that using more stratifiers can be beneficial, especially in reducing the variances in the size measure among substrata.

Variation among the stratification scheme results

One of the objectives of this paper was to determine if the efforts such as those undertaken by Kish and cohorts was worth the effort. Table 7 shows the 10^{th}, 50^{th} and 90^{th} percentiles of the distribution of the between PSU variance and equal size strata measure across all substratification schemes in each major stratum. The table shows much more variation among the equal size strata measure than the between PSU variance. This is also seen in the scatterplots for each MSA major stratum in Figure 2 and each Non-MSA major stratum in Figure 3. The objective is to find the point that is furthest in the lower left-hand corner of each plot. Certainly, efforts to identify key stratifiers that are highly correlated with survey outcome measures

will help to reduce the between PSU variance, in some major strata more than others, and one could search and select a scheme that approaches optimality. Also, there remains the benefit of reducing the equal size strata measure. Since there is a lot of variation between stratification schemes in terms of the between PSU variance and the equal size strata measure, we can say that it is well worth the effort to evaluate several stratification schemes. However, key auxiliary data are needed at the time of stratification to reduce the between PSU variance.

Table 6. Percent relative reductions between the resulting minimum evaluation measures when SV=2 and when SV=4

Census Division	MSA Status	Number of schemes		Difference between the resulting minimum values (*SV*=2 – *SV*=4)	
		SV=2	*SV*=4	Between PSU variance	Equal size strata measure
1	MSA	2	12	-2.2%	-67.7%
1+2	Non-MSA	2	4	0.0%	-93.8%
2	MSA	64	4015	-6.4%	-15.0%
3	Non-MSA	4	16	0.0%	-66.8%
3	MSA	1680	--	--	--
4	Non-MSA	4	16	0.0%	-77.1%
4	MSA	8	64	-21.4%	-27.0%
5	Non-MSA	8	64	-16.0%	-47.6%
5	MSA	723	--	--	--
6	Non-MSA	4	16	-20.8%	-62.8%
6	MSA	8	64	-23.0%	-37.9%
7	Non-MSA	4	16	-11.7%	0.0%
7	MSA	32	1024	-20.3%	-50.0%
8	Non-MSA	2	4	-1.1%	0.0%
8	MSA	8	64	-1.1%	0.0%
9	Non-MSA	2	4	-0.4%	0.0%
9	MSA	128	16269	-20.4%	-12.1%

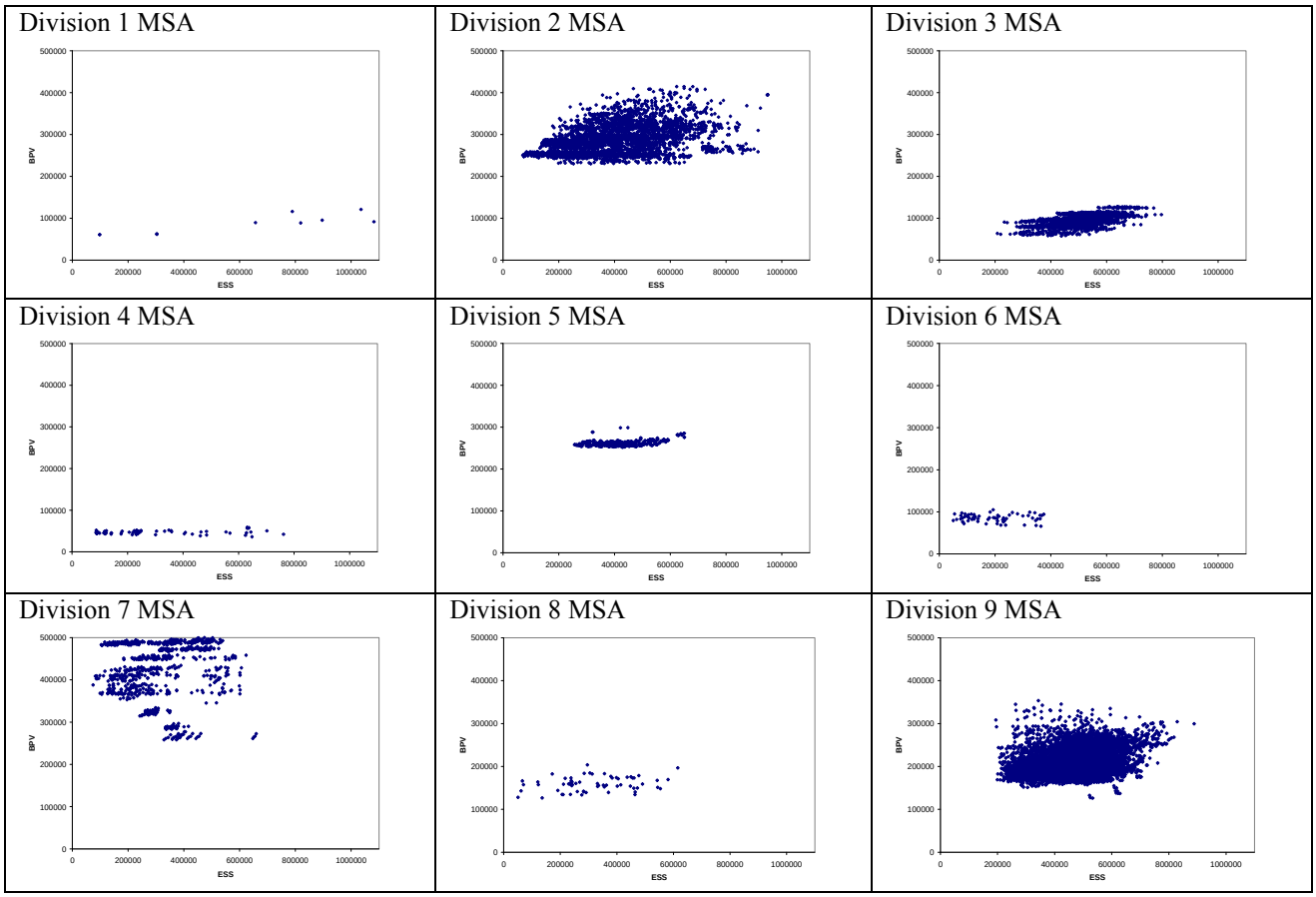
Note: For MSAs in divisions 3 and 5, only schemes with *SV*=2 were generated.

Table 7. Distribution of between PSU variance and equal size strata measure by major strata for the evaluation runs

Census Division	MSA Status	Between PSU variance			Equal size strata		
		10^{th} percentile	Median	90^{th} percentile	10^{th} percentile	Median	90^{th} percentile
1	MSA	60,955	87,286	116,283	97,979	723,041	1,081,964
1+2	Non-MSA	73,237	101,596	114,776	891	9,169	32,104
2	MSA	249,163	280,849	332,543	190,117	392,078	614,662
3	Non-MSA	44,928	61,189	63,345	4,038	15,871	21,743
3	MSA	75,963	94,687	111,795	376,108	506,648	626,763
4	Non-MSA	94,137	99,486	102,205	4,888	13,844	25,130
4	MSA	41,655	46,791	51,856	96,229	237,762	631,251
5	Non-MSA	176,469	200,236	219,718	7,766	22,962	43,547
5	MSA	255,651	260,085	266,973	324,870	411,910	525,214
6	Non-MSA	108,569	121,440	134,659	5,713	17,779	32,291
6	MSA	71,767	85,535	95,401	83,546	187,780	347,391
7	Non-MSA	223,732	249,608	264,046	5,631	12,213	20,344
7	MSA	323,910	455,321	490,038	158,548	322,551	482,599
8	Non-MSA	134,652	139,215	151,341	3,283	8,686	29,101
8	MSA	134,808	158,567	182,306	135,773	300,397	477,877
9	Non-MSA	155,730	159,306	168,099	4,013	7,602	36,951
9	MSA	181,132	203,387	238,806	317,674	458,676	574,814

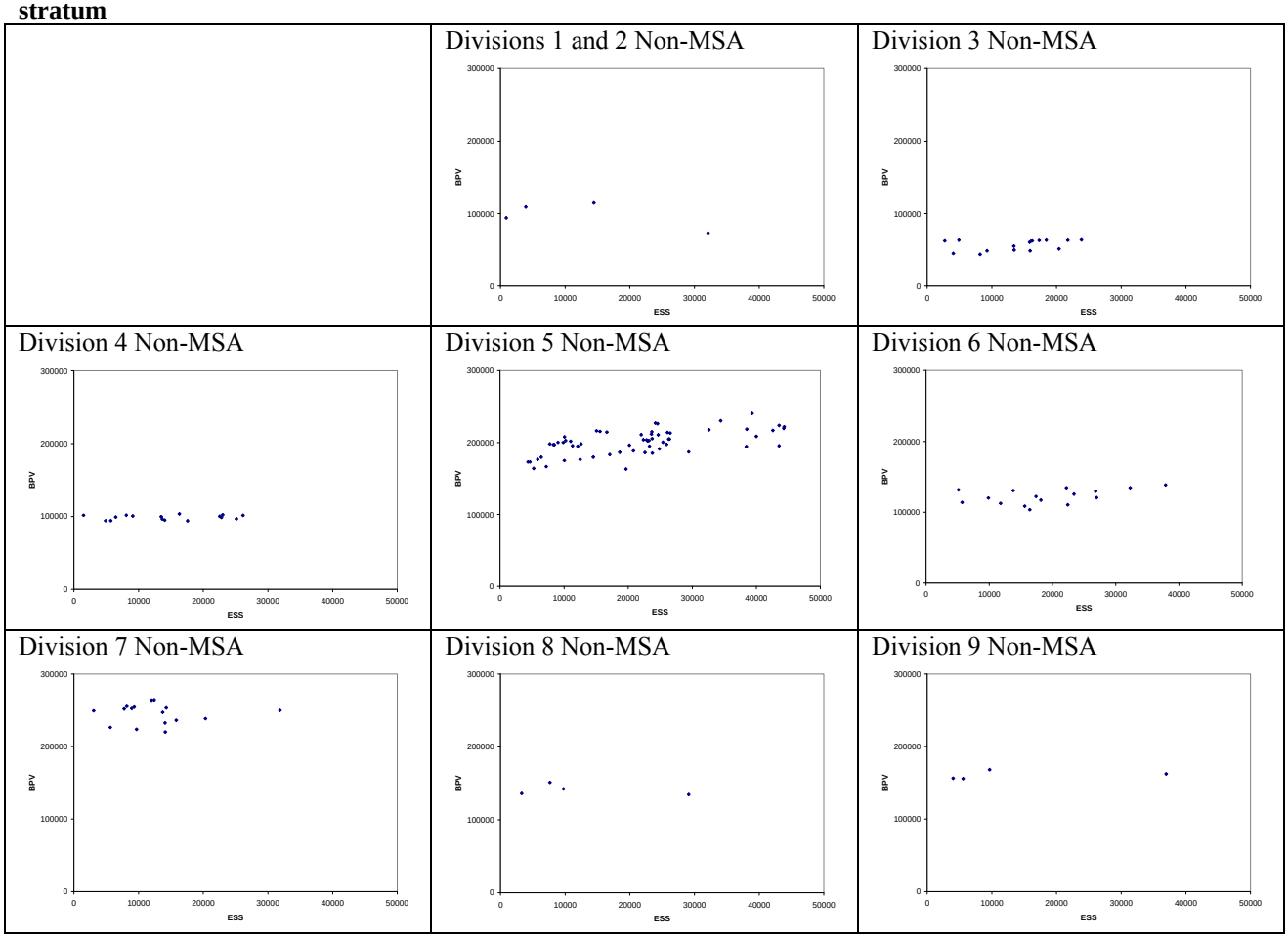
Note: Square roots of each measure are shown.

Figure 2. Scatterplots showing the between PSU variance and equal size strata measure for each MSA major stratum



Notes: The x-axis is the equal size strata measure (ESS) and the y-axis is the between PSU variance (BPV)

Figure 3. Scatterplots showing the between PSU variance and equal size strata measure for each Non-MSA major stratum



Notes: The x-axis is the equal size strata measure (ESS) and the y-axis is the between PSU variance (BPV)

As mentioned above, we are interested in how much better the 'best' evaluation scheme is over the 2003 NAAL scheme. Figure 4 compares the NAAL scheme, depicted as a dot, and the associated best evaluation scheme, depicted by the arrow. Each line is for a comparison between the NAAL scheme and the best evaluation scheme for a certain census division. This is for MSAs only. The x-axis is the equal size strata measure and the y-axis is the between PSU variance. The objective is to reduce the values in each axis and therefore head into the lower left hand corner. The best evaluation scheme was determined by the average combined rank. As you can see, some of the lines point down to the lower left corner, showing the improvement can be made to both the equal size strata measure, and the between PSU variance. However, some point down to the right, showing improvement in between PSU variance but not in equal size strata measure. In Division 2, the equal size measure was reduced substantially at the expense of some slight increase to the between PSU variance. While this was a tough test for the NAAL scheme, we were moderately satisfied with the results, while there is still room for improvement.

Figure 4. Comparison of the NAAL scheme and the best evaluation scheme, by Census Division, MSAs only

Concluding remarks

At the time of the 2003 NAAL stratification process, there was no measure of the association between the stratifiers and key survey outcome variables. With the production of county-level indirect estimates of the percent lacking *BPLS*, an evaluation variable that is one of the key NAAL survey outcomes became available, and key predictor variables were identified for the SAE model. Much of the effort, once software such as the one developed here is established, should go into the identification of key stratifiers. Soon, sample design plans will be written for the 2011 Programme for International Assessment of Adult Competencies (PIAAC). The PIAAC survey is similar to the 2003 NAAL in that it measures outcomes related to literacy through an in-person assessment. As we have seen from the evaluation, the use of the key predictors in the SAE process as stratifiers will help reduce the between PSU variance in the PIAAC survey. That is, the plan will include the percent lacking *BPLS* as the evaluation variable, while forming explicit strata using demographic data (the SAE predictors) from the most recent decennial Census. Also, when using the underlying approach, the use of more stratifiers when creating the stratification schemes will help to reduce the equal size strata measure.

In Kish (1965) page 379, he says "A great many man-hours were spent in the stratification process. However, it is questionable whether the amount of time devoted to reviews and refinements paid off in appreciable reductions in sampling variances. Intuitive notions about gains from stratification can be misleading." As a result of the evaluation of the 2003 NAAL stratification process, we have, in effect, revisited the efforts conducted by Kish and his colleagues, by measuring the variation across numerous stratification schemes, using a computer-intensive search. Under the nested stratification approach described in this paper, which arrives at explicitly defined boundaries, we found that there is considerable variation among resulting schemes in terms of the equal size strata measure. Reducing this component will lead to equalizing interviewer workloads and reducing the variation in estimated totals and the bias of these variance estimates -- not so much the variation in proportions and means. We found less variation in most major strata with regards to the between PSU variance, given a strong set of stratifiers. However, some major strata experienced considerable between PSU variance, for which reduction in sampling variances can be realized. Therefore, we conclude that benefits can be realized by using a constructive, systematic, and thorough approach for identifying a stratification scheme for PSUs. We also recommend the development of software to generate many stratification schemes to facilitate an analysis of the many solutions. Lastly, we recommend that repeating surveys use information from the prior round to improve the stratification process. With regards to future research, it would be interesting to compare results from the simplistic and efficient nested approach to the more sophisticated computer-intensive clustering algorithms, while weighing in the level of effort involved.

References

Friedman, H. P. and Rubin, J. (1967). On some invariant criteria for grouping data. *Journal of the American Statistical Association*. Vol. 62, 1159-1178.

Jewett, R.S. and Judkins, J. (1988). Multivariate stratification with size constraints. SIAM Journal of Scientific Statistical Computing, Vol. 9, No. 6, 1091-1096.

Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.

Ludington, P. (1992). Stratification of primary sampling units for the Current Population Survey using computer intensive methods. *Proceedings of the Section on Survey Research Methods of the American Statistical Association.*

Mohadjer, L., Kalton, G., Krenzke, T., Liu, B., Van de Kerckhove, W., Li, L., Sherman, D., Dillman, J., Rao, J. and White, S. (2009). National Assessment of Adult Literacy: Indirect county and state estimates of the percentage of adults at the lowest level of literacy for 1992 and 2003 (NCES 2009-482). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Rao, J.N.K. (2003). *Small area estimation*. Wiley-Interscience.

Appendix A

The following discusses the substratification approach in more detail. Suppose a is the 1st stratifier, b the 2nd stratifier and c the 3rd stratifier (if $SV = 3$). Then (a_1, b_1, c_1) defines substratum 1 (i.e., 'ending' node), (a_2, b_2, c_2) defines substratum 2, …, (a_H, b_H , c_H) defines substratum H. To define the tree-structure given H_g and SV (for this explanation we set $SV = 3$) the following constraints are used:

1. $a_i + b_i + c_i < H_g + SV$ for any set of (a_i, b_i, c_i) defining node i.

2. The set of ending nodes $I = 1, 2, \ldots, H_g$ defining a set of substrata must always contain ending node (1,1,1).

3. There are no gaps between a_i and a_j, or b_i and b_j, or c_i and c_j, for all i and j that comprise the ending nodes defining a set of substrata.

4, Similar to 3), for a given value for stratifier a, there must exist a sequence of values (or one value) starting at 1 for b_i, for all i relating to the given value for stratifier a. In the same manner, for a given value for stratifier b, there must exist a sequence of values (or one value) starting at 1 for c_i, for all i relating to the given value for stratifier b.

The sets of all possible substrata (or ending nodes) are the combinations of stratifiers a, b, and c, where $a = 1, 2, \ldots, H_g$; $b = 1, 2, \ldots, H_g$; $c = 1, 2, \ldots, H_g$, that satisfy the above constraints. At this point, there is a tree structure defined for each possible substratification scheme for a major stratum. However, the branches have not been explicitly defined within each of the trees. To make branches from a node, it does the following:

1. Counts the number of branches formed by the first stratifier. Let us call it A.

2. Counts the number of substrata, holding the value of the stratifier constant. That is, for the split on the first stratifier a, count the number of ending nodes that result from $a = 1$ and call it $H_{a=1}$. Do that for each value of a that results from the first stratifier to arrive at $H_{a=1}$, $H_{a=2}$, … $H_{a=A}$.

3. Sorts by the stratifier.

4. Creates A–1 cutpoints on the stratifier a. The 1st cutpoint is the stratifier a value contributing $100\%*H_{a=1}/H_g$ of the total measure of size for the subpopulation defined by the particular node (which is the major stratum if it is the 1st stratifier). The 2nd cutpoint is the stratifier a value defining $100\%*(H_{a=1} + H_{a=2}) / H_g$ of the total measure of sizes for the subpopulation defined by the particular node, and so on.

5. For the second stratifier b, analogously repeats steps 1)-3) for each non-ending node created by stratifier a. In the same manner, continue with stratifier c.

As an example, suppose $H_g = 4$ and $SV=4$. Let the scheme be (1,1,1,1) (1,1,2,1) (1,1,3,1) (2,1,1,1). For stratifier a, $H_a=(1) = 3$ and $H_a =(2) = 1$. Therefore, the percentile cutoff is 75% for first branch and 25% for the second branch. For stratifier b: $H_b=(1,1) = 3$; $H_b =(1,2) = 1$. But since the parent $H_a =(1)$ has just 1 immediate child, and parent $H_a =(2)$ has just 1 immediate child, the number of cutpoints = 0 for each parent branch, and so we don't need to compute the cutpoints using stratifier b. For stratifier c, $H_c=(1,1,1)$ has 1 ending node, $H_c =(1,1,2)$ has 1 ending node, and $H_c =(1,1,3)$ has 1 ending node. The parent $H_b =(1,1)$ has 3 immediate children and therefore 2 cutpoints are made. The cutpoints use the number of ending nodes from the parent H_b (1,1), which is equal to 3 (in general it does not equal the number of immediate children, but it does in this example). So $100\%* H_c =(1,1,1) / H_b = (1,1)$ and $100\%*(H_c =(1,1,1) + H_c =(1,1,2)) / H_b = (1,1)$ or 33.3% and 66.6%. And since parent $H_b =(1,2)$ has 1 immediate child, there is no cutpoint formed. For stratifier d, each parent H_c (1,1,1), H_c (1,1,2), H_c (1,1,3) and H_c (2,1,1) are all equal to 1 so there are no cutpoints generated for stratifier d.