A New Approach for Multiway Stratification: School Sampling in Charting the Progress of Education Reform: An Evaluation of the Recovery Act's Role

Lou Rizzo[1], Jane Li[1], Yong Lee[1]
Babette Gutmann[1], Patty Troppe[1]
[1]Westat, 1600 Research Blvd, Rockville, MD 20850

Abstract
Charting the Progress of Education Reform: An Evaluation of the Recovery Act's Role is sponsored by the Institute for Education Sciences in the U.S. Department of Education. The evaluation assesses how states, districts, and schools are working to implement education reforms. We required a nationally representative sample of school districts and schools to examine the role that Recovery Act programs may have played in such efforts. The school sample was nested within the district sample (a two-stage design), and we required at least two sampled schools within each district for analysis purposes. In addition, we required stratification control for grade span and school performance. This required then a multiway (three-way) stratification structure. To carry this out, we utilized the new balanced sampling theory as developed by Deville and Tillé (e.g., Deville and Tillé (2004)). This paper describes our new methodology for executing this theoretical approach, given the large sample sizes, and also presents evidence that the methodology was successful in meeting the desired criteria (respecting the desired probabilities of selection, meeting the three sets of stratification criteria, and being a randomized rather than a controlled sampling approach).

Key Words: balance sampling, multiway stratification, school sampling, martingale

1. The Problem

The *Charting the Progress of Education Reform: An Evaluation of the Recovery Act's Role* survey is sponsored by the Institute for Educational Sciences in the US Department of Education, and studies the role of funding in the development of reform efforts in the national educational system. A major focus is the effect of the 2009 American Recovery and Reinvestment Act allocations to the educational sector (e.g., the Race To The Top). As part of this effort, a nationally representative sample of districts and schools was required.

A total of 1,700 districts were included in this sample from the 2009-2010 Common Core of Data (CCD) public school district frame. These districts were sampled probability proportionate to enrollment, with high-poverty districts[1] oversampled by a factor of 2.75. There were 378 districts which had measures of size exceeding the sampling interval, and these became 'certainty districts' (in the sample with probability 1). Information is being canvassed from these 1,700 sampled districts regarding their reform agendas and how they are implementing reform programs in the years 2011 through 2014.

Part of the study also included gathering of information at the school level to check on whether or not the schools were conforming with district reform programs. The sample sizes necessary for this purpose were deemed to be small, but it was considered important ***to have at least two sampled schools per sampled district*** (except for districts with only one school, in which case the single school was selected with certainty). Only large certainty districts had school sample sizes exceeding two. A total of 3,800 schools were sampled, nested within the 1,700 sampled districts (a mean sample size of 2.24 sampled schools per sampled district).

After removing schools in one- and two-school sampled districts (all schools were selected with certainty within the sampled district in these two cases), the school sampling process was probability proportionate to school enrollment, with assigned school samples for most districts equal to 2, except for the larger certainty districts. The districts were a stratification variable for school sampling, in fact a "deep" stratification (small sample sizes per stratum). If districts were the only stratification variable, then this sampling process would essentially be two-stage sampling, with districts the first stage and schools the second stage of sampling within distrits. But, it was desirable also to stratify the schools according to school span (elementary, middle, high, and other), to control the sample sizes for these four classes of schools, and school size within school span, and also to stratify the schools according to 'school performance' (Persistently Low Achieving (PLA) schools, non PLA SINI schools (Schools in Need of Improvement, non SINI schools). PLA schools were oversampled by a factor of 8 given their great importance in this study. In addition, school performance needed to be stratified by school span, as high schools tended to be PLA and SINI schools more frequently, and elementary schools less frequently. To have these extra stratification structures in addition to the district stratification precluded any simple two-stage sample design, as there was no room for strata within each district sample. We required a multiway stratification scheme: in particular, a three-way stratification structure.

2. Potential Solutions

Multiway stratification and its allied procedure controlled selection has a long history in survey sampling. The original paper introducing controlled selection was Goodman and Kish 1950. Bryant et al. 1960 presents a scheme for carrying out two-way stratification, which Jessen 1970 extends to three-way stratification. Causey et al. 1985 and Sitter and Skinner 1994 present further methodologies for carrying out controlled selection to implement multiway stratification. These methodologies have the benefits of matching desired inclusion probabilities exactly, but require extensive, intricate linear

[1] These were districts with more than 21.66% of their children in families below the poverty line, according to SAIPE (Small Area Income and Poverty Estimates program of the Bureau of the Census. 21.66% was chosen as the weighted 75th percentile: i.e., exactly 25% of the district enrollment was in districts exceeding this cutoff percentage). Note that some districts such as charter school districts were not in SAIPE and had poverty percentages imputed for them based on geographic location.

programming efforts to ensure this. These methodologies are nearly impossible to implement when the sample sizes are greater than 3,000, with more than 1,000 strata in one stratification dimension, as all of these authors admitted. In addition, variance estimation is a difficult issue.

A new methodology that can be utilized to implement multiway stratification is cube sampling (Deville and Tillé 2004). The 'flight phase' of cube sampling operates by defining a stochastic process with T vectors of length N. Each entry in each vector corresponds to a frame element, so that N is equal to the frame size. Each vector corresponds to one sample unit being sampled, so that T is equal to the flight phase sample size. The first vector in the stochastic process (before the first sample unit is sampled) will have the probabilities of selection for each frame unit. The next vector will replace these probabilities with a '1' for one unit (which becomes the first sampled unit), with a '0' for one or more units (which become definitively not sampled), and with revised conditional probabilities for all other frame units. If the flight phase completes the sampling process the final vector will consist of only 1's and 0's, with the frame units with a '1' being the sampled elements and the frame units with a '0' being the unsampled elements. This vector will simply present the final sample. If the flight phase does not complete the sampling process, then the final vector will consist of a mixture of 1's and 0's and conditional probabilities. The 1's correspond then to elements definitively sampled, the 0's to elements definitively not sampled, and the conditional probabilities for the remaining elements are the current conditional probabilities of selection. The 'landing phase' then completes the sampling process that the flight phase left unsampled, starting with this final vector as a starting point.

The vectors in the stochastic process form a martingale process, i.e., the expected values of each vector are equal to the previous vector. All vectors have the same unconditional expectation, equal to the first vector of probabilities of selection. It is this property which maintains the desired probabilities of selection, while still allowing for the constraints coming from the multiway stratification process.

Cube sampling achieves the sample design desired, and in addition there are approximate consistent variance estimators available (see Deville and Tillé 2005). A drawback of the method is that the flight phase may sometimes end too soon, requiring too much to be sampled through the landing phase. The landing phase is unwieldy if there is too much sample size left to select[2]. This in fact occurred in our application.

Falorsi and Righi (2008) present an application of cube sampling to carry out multi-way stratification in the context of small area estimation. Our approach has some overlap with their approach as is pointed out below.

3. The Sequential Raked Balanced Sampling Solution

The sequential raked balanced sampling solution is based on the cubed sampling approach of Deville and Tillé 2004, but with some important modifications. First, it should be noted that of the original 1,700 sampled districts, 107 had only one or two

[2] The public use software available for cube sampling (the macro **samplecube** in http://www.cran.r-project.org/web/packages/sampling) cannot go beyond relatively small sample sizes in the implementation of the landing phase (as all possible samples are enumerated).

schools. All schools were sampled within these districts, and these were set aside. This leaves a frame of 38,736 schools within 1,593 sampled districts. The school sample size is 3,608 (the designated sample size of 3,800 minus the 192 schools in the one- and two-school districts).

The school frames are divided into eight separate frames based on three dichotomous indicators: district poverty status (whether the district is high-poverty[3] or not), district certainty status (whether or not the district was a certainty district in district sampling), and whether or not the district has at least one Persistently Low Achieving (PLA) school. Balanced sampling is carried out separately within each of these eight major strata defined at the district level.

The first step is to define unconditional probabilities of selection for the process. These will reflect the overall probability of entering the sample at the end of the process, and are then used in developing the sampling weights (the inverse of these unconditional probabilities of selection multiplied to the inverse district probabilities of selection are the initial school base weights). We defined desired sample sizes for the three stratification margins by adding together measures of size in each stratification cell and rounding these to integer values[4]. Then, the preliminary measures of size were raked to these integer totals. The probabilities were raked so that they added to the rounded sample sizes for each cell in each of the stratification margins. It should be noted that this ‘distorts’ the probabilities based on the original measures of size so that they fit the rounded, adjusted totals. Note that this approach was utilized by Falorsi and Righi (2008): see their Section 3.2.

Table 3-1 provides frame sizes and sample sizes for the final school frame, as well as the coefficient of variation of the ratio between the final raked weight and the original weight proportional to the initial measure of size. Also included is 1 plus the CV squared, a measure of a design effect induced from a ‘haphazard’ variability in the weights (see for example Kish 1992). This is a measure of the effect of the ‘distortion’ from adjusting the probabilities to agree with rounded, consistent sample sizes for each of the three stratification dimensions, and reflects the increase in variability induced from corseting the probabilities in this way. It is hoped that the gains in precision from having the multiway stratification offset this loss: if not, the multiway stratification is not justified.

---

[3] High-poverty is defined as greater than the 75th percentile for the percentage of families in poverty (as per the SAIPE estimates provided by the US Bureau of the Census).

[4] There was also an extra step to adjust the values so that matching totals across the three stratification margins were equal.

Table 3-1: Final School Frame Sample Sizes and CVs of Weight Ratios

Poverty stratum	*District certainty stratum*	*PLA status*	*School frame size*	*School sample size*	*CV of weight ratios (times 100)*	*1+CV²*
High	*Noncrt*	*noPLA*	2700	796	9.48	1.009
High	*Noncrt*	*w/PLA*	819	191	49.53	1.245
High	*Cert*	*noPLA*	2782	269	49.15	1.242
High	*Cert*	*w/PLA*	9713	621	31.31	1.098
Low	*Noncrt*	*noPLA*	8825	1,270	1.21	1.000
Low	*Noncrt*	*w/PLA*	2557	179	45.09	1.203
Low	*Cert*	*noPLA*	4518	130	36.62	1.134
Low	*Cert*	*w/PLA*	6822	152	35.20	1.124
All	*All*	*All*	38,736	3,608		

The balanced sampling draws 3,608 noncertainty schools within 1,593 districts, done separately within each of the eight major strata. For simplicity's sake, we will not include in this report the s subscript for the major strata. It should be understood that all formulas and calculations below refer to each major stratum one by one. The object of this balanced sampling was to select the schools respecting the final measures of size $\breve{p}_{ij}$ (district i, school j) while simultaneously respecting three sets of assigned stratification sample sizes:

- $n_i, i = 1, \ldots, I,$ where I is the total number of districts;
- $n_a, a = 1, \ldots, A,$ where a corresponds to a span-performance stratum cell, and A is the total number of span-performance strata cells.
- $n_b, b = 1, \ldots, B,$ where b corresponds to a span-size stratum cell, and B is the total number of span-performance strata cells.

The batches are designated in order as $c = 1, \ldots, C$. The districts within each batch t are designated as $\mathcal{D}_c$, with these districts indexed as $i = 1, \ldots, I_c$. Schools within districts are indexed as $iabj$, with i representing district, $a = 1, \ldots, A$ representing the span-performance stratum containing the school, $b = 1, \ldots, B$ representing the span-size stratum, and $j = 1, \ldots, S_{iab}$ representing the specific school within the cell defined by district i, and strata a and b. S_{iab} is the number of schools in each of these district-strata cross-cells. It should be noted that some of the possible cross-cells are empty: not every district has every a- b cross cell represented within it. Each iteration of the sequential process is designated as $t = 1, \ldots, T$, with each iteration having two components : 1—completing the sampling for one batch, and 2—updating the conditional probabilities in preparation for the next iteration. Thus, iteration t=2 will complete the sampling for batch c=2, and update the probabilities in preparation for the sampling of batch c =3. Note that C is equal to T (the number of batches is equal to the number of iterations), but we keep the notation separate for clarity. We also have a t=0 which represents the values of particular parameters before the first batch is sampled: the 'initial values' of these parameters.

Following the general concept of Deville and Tillé (2004) in defining a discrete time stochastic process define a set of vectors $\boldsymbol{\pi}(t)$, $t = 0, \ldots, T$, for each time t, with the first vector $\boldsymbol{\pi}(0)$ consisting of the initial inclusion probabilities for each school $\breve{p}_{ciabj}$, ordered by batch, district, stratum a, stratum b, and school with cells $ciab$:

$\boldsymbol{\pi}(0) = \{\boldsymbol{\pi_1}(0), \dots, \boldsymbol{\pi_c}(0), \dots., \boldsymbol{\pi_C}(0)\}$ \
with each $\boldsymbol{\pi_c}(0) = \{\boldsymbol{\pi_{c1}}(0), \dots, \boldsymbol{\pi_{ci}}(0), \dots., \boldsymbol{\pi_{cI_c}}(0)\}$
and each $\boldsymbol{\pi_{ci}}(0) = \{\breve{p}_{ci111}, \dots, \breve{p}_{ciabj}, \dots., \breve{p}_{ciABS_{ciAB}}\}$

The vectors $\boldsymbol{\pi_{ci}}(0)$ can also be written as

$\boldsymbol{\pi_{ci}}(0) = \{\tilde{p}_{ci111}(0), \dots, \tilde{p}_{ciabj}(0), \dots., \tilde{p}_{ciABS_{ciAB}}(0)\}$

The initial probability $\tilde{p}_{ciabj}(0)$ is equal to the unconditional probability $\breve{p}_{ciabj}$.

Each vector $\boldsymbol{\pi}(t)$ following $\boldsymbol{\pi}(0)$ represents the situation after sampling one batch fully, and then revising (via raking) the remaining probabilities to satisfy the balancing equations as they are after finishing all batches up to that point. Let's move to the second vector $\boldsymbol{\pi}(1)$, which represents the situation after completing the selection of batch 1 and re-raking the conditional probabilities. The subvectors of $\boldsymbol{\pi}(1)$ are defined as follows. The first subvector $\boldsymbol{\pi_1}(1)$ corresponding to the first batch which is sampled out in step 1 is equal to a vector of 1's and 0's (called $\mathbf{I_1}$) with the 1's representing the sampled schools, and the 0's the nonsampled schools. All other subvectors $\boldsymbol{\pi_c}(1), c = 2, \dots, C$ in $\boldsymbol{\pi}(1)$ are equal to the revised set of conditional probabilities $\tilde{p}_{ciabj}(1)$ (revised to be consistent with the revised sample sizes following the first batch selection).

$\boldsymbol{\pi}(1) = \{\mathbf{I_1}, \boldsymbol{\pi_2}(1) \dots, \boldsymbol{\pi_c}(1), \dots., \boldsymbol{\pi_C}(1)\}$
with each $\boldsymbol{\pi_c}(1) = \{\boldsymbol{\pi_{c1}}(1), \dots, \boldsymbol{\pi_{ci}}(1), \dots., \boldsymbol{\pi_{cI_c}}(1)\}, c = 2, \dots, C$
with each $\boldsymbol{\pi_{ci}}(0) = \{\tilde{p}_{ci111}(1), \dots, \tilde{p}_{ciabj}(1), \dots., \tilde{p}_{ciABS_{ciAB}}(1)\}, \ c = 2, \dots, C$

For the general step t, $\boldsymbol{\pi}(t)$ is equal to

$$\boldsymbol{\pi}(t) = \{\mathbf{I_1}, \dots, \mathbf{I_t}, \boldsymbol{\pi_{t+1}}(t), \dots., \boldsymbol{\pi_C}(t)\}$$

where $\mathbf{I_1}$ through $\mathbf{I_t}$ consist of 0's and 1's representing the final samples for batches 1 through t, and $\boldsymbol{\pi_{t+1}}(\mathrm{t})$ through $\boldsymbol{\pi_C}(t)$ are equal to revised sets of conditional probabilities following the selection of the school sample from batch t.

The final vector $\boldsymbol{\pi}(T)$ consists of $\mathbf{I_1}$ through $\mathbf{I_C}$ (only 0's and 1's) and is the final vector of school sample selections.

Deville and Tillé's theory envisages that the $\boldsymbol{\pi}(t)$, $t = 1, .., T$ will be a martingale process, i.e., all of the $\boldsymbol{\pi}(t)$ vectors will have conditional expectation equal to the preceding vector $\boldsymbol{\pi}(t-1)$, while also satisfying the balancing conditions (see Definition 5 in Deville and Tillé (2004). For the schools that are subject to being sampled at time t, this will mean that the sample indicator vector of 1's and 0's will have as its probabilities the conditional probabilities from $\boldsymbol{\pi}(t-1)$. These conditional probabilities then have expectation equal to the preceding set of conditional probabilities, and so on back to the set of unconditional probabilities in $\boldsymbol{\pi}(0)$.

We can facilitate things by defining the following random variable as is done in Deville and Tillé (2004):

$$\pi_{ciabj}(\mathrm{t})=\begin{cases}\tilde{p}_{ciabj}(t) & t < c\\ 0 \text{ or } 1 & t \geq c\end{cases}$$

With this definition, each $\boldsymbol{\pi}(t)$ is simply a vector of the random variables $\pi_{ciabj}(t)$ in the appropriate ordering.

The next random variables we will define are summations of the $\pi_{ciabj}(t)$. We can prove certain summations do in fact have the martingale property. For example, suppose we sum over the schools within each district ci. These will add to the district sample size $n_{ci\ldots}$ for each and every iteration t:

$$n_{ci\ldots} = \sum_{a=1}^{A}\sum_{b=1}^{B}\sum_{j=1}^{S_{ciab}} \pi_{ciabj}(t)$$

For $t < c$ (before the district is selected out) the formula expresses that the summation of the $\tilde{p}_{ciabj}(t)$ equals $n_{ci\ldots}$. This in fact is 'guaranteed' by the raking process, which has district as a dimension. For $t \geq c$ (after the district is selected out) the formula expresses that the summation of the sample indicators in the district is equal to $n_{ci\ldots}$ (i.e., that $n_{ci\ldots}$ schools were selected within district ci). It should be noted that these summations of random variables are in fact a martingale process for t=1,.., T: the summations are equal to a common expectation $n_{ci\ldots}$ with certainty.

Suppose now we sum over the schools within each span-performance stratum cell a. These will add to the stratum cell sample size $n_{..a..}$ for each and every iteration t:

$$n_{..a..} = \sum_{c=1}^{C}\sum_{i=1}^{I_c}\sum_{b=1}^{B}\sum_{j=1}^{S_{ciab}} \pi_{ciabj}(t)$$

This summation of random variables again will add to a fixed value for all t=1,…,T. In this case (unlike the district case), each and every t will have $\pi_{ciabj}(t)$ that are 0's and 1's and $\pi_{ciabj}(t)$ that are conditional probabilities, as we are summing over all batches (except for t=0, in which case all of the $\pi_{ciabj}(t)$ are probabilities, and for t=T, in which case all of the $\pi_{ciabj}(t)$ are sample indicators). Again the fact that this summation of random variables adds to a fixed value is guaranteed by the raking process, which rakes the $\pi_{ciabj}(t)$ to equal this fixed value within each a-cell. This summation is a martingale process again, in a trivial sense (all conditional expectations are equal to the fixed expected value $n_{..a..}$).

In a similar way, we can sum over the schools within each span-performance stratum cell b. These will add to the stratum cell sample size $n_{...b.}$ for each and every iteration t:

$$n_{....b.} = \sum_{c=1}^{C}\sum_{i=1}^{I_c}\sum_{a=1}^{A}\sum_{j=1}^{S_{ciab}} \pi_{ciabj}(t)$$

This summation of these random variables is a martingale process.

Another martingale process that is a function of the $\pi_{ciabj}(t)$ random variables are the proportions $p_j^{(ciab)}(t)$ within each nonempty cell $ciab$ (district crossed with span-performance and span-size strata cells):

$$p_j^{(ciab)}(t) = \frac{\pi_{ciabj}(t)}{\sum_{j=1}^{S_{ciab}} \pi_{ciabj}(t)}$$

For $t < c$ (before the district is selected out) the $p_j^{(ciab)}(t)$ are equal to the proportions that each school has of the overall aggregate measure of size in the cross-cell. The raking process guarantees that these cross-cell proportions do not change as the $\pi_{ciabj}(t)$ change as t increases: raking only changes the total measure assigned to every cross-cell. This is a well-known property of raking (all observations in the lowest-level cross-cells receive the same raking adjustment). For $t \geq c$ (after the district is selected out) the $p_j^{(ciab)}(t)$ will be equal to 1 for one school and 0 for all other schools if one school is sampled, will be equal to ½ for each sampled school if two schools are sampled, and 0 for all other schools, etc.

Another important stochastic process that is a summation of the $\pi_{ciabj}(t)$ is the summation over the district-span performance-span size cross cells:

$$n_{ciab}(t) = \sum_{j=1}^{S_{ciab}} \pi_{ciabj}(t)$$

We cannot prove that each and every $n_{ciab}(t)$ by itself is a martingale process, in other words that for every nonempty $ciab$ and every $n_{ciab}(t)$ we have $E\big(n_{ciab}(t)\big)$=$n_{ciab}(t-1)$. In that sense, this procedure does not satisfy the balanced sampling definition (Definition 5 in Deville and Tillé (2004)). But we ***can*** say that particular linear combinations are martingales, in particular

$$\sum_{a=1}^{A} \sum_{b=1}^{B} n_{ciab}(t) = n_{ci\ldots}$$

$$\sum_{c=1}^{C} \sum_{i=1}^{I_c} \sum_{b=1}^{B} n_{ciab}(t) = n_{..a..}$$

$$\sum_{c=1}^{C} \sum_{i=1}^{I_c} \sum_{a=1}^{A} n_{ciab}(t) = n_{\ldots.b.}$$

All of these linear combinations are martingales, since they add to fixed values for each t=1,…, T. This controls the degree of systematic deviation of the $n_{ciab}(t)$ from their predecessor values $n_{ciab}(t-1)$. If the conditional expectation of, say, $n_{ciab}(t)$ exceeds $n_{ciab}(t-1)$, then another cell, say $n_{cia'b}(t)$, must have its conditional expectation less than $n_{cia'b}(t)$ to allow equality of the marginal totals by district, a-cell, and b-cell for every t=1,…, T. In this sense the sequence of conditional expectations of the cross-cells is well-controlled. But exact martingale properties for each and every cross-cell cannot be asserted.

4. Empirical Evaluation of the Sequential Raked Balanced Sampling Solution

We evaluated the balanced sampling algorithm by running it 100 times independently. This simulation study allows us to generate mean-squared errors for the sample sizes for the span-performance and span-stratum sample sizes. These will be used in the process of generating variance estimators. The simulation study also allows us to evaluate the inclusion probabilities for each school on the frame empirically.

Table 4-1 presents the variability in the final sample sizes for each span-performance stratum cell over these 100 samples. The summary values are the simulation mean (the mean sample size over the 100 samples), the simulation standard deviation (the standard deviation of sample sizes), and the simulation root mean squared error (the mean deviation of the sample sizes from the target value). Table 4-2 provides similar calculations for the span-size stratum cells.

Table 4-3 presents data on the true inclusion probabilities for each school on the final frame. Each row corresponds to a set of frame schools with similar assigned probabilities of selection (a 'bin'). For example, the first row represents frame schools with assigned probabilities between 0 and 0.5%. Given are the minimum, mean, and maximum probabilities, the aggregate measure of size (which is also the expected sample size), the mean value of sample indicators of all schools on the frame in this 'bin', and the mean sample percentage. The mean value of sample indicators should have as its expected value the aggregate measure of size, and the mean sample percentage should have as its expected value the mean bin probability. This does appear to be the case.

Table 4-1: Span-performance Strata Cells with Balance Results from 100-Sample Simulation

Poverty stratum	*District certainty stratum*	*PLA status*	*Span*	*PLA/SINI status*	*Assigned sample size*	*Mean*	*Std Dev*	*Root MSE*
High	*Noncrt*	*NoPLA*	*Elem*	*NonPLA*	283	282.9	0.80	0.80
High	*Noncrt*	*NoPLA*	*Elem*	*regSINI*	76	76.0	0.68	0.68
High	*Noncrt*	*NoPLA*	*High*	*NonPLA*	175	175.0	0.92	0.92
High	*Noncrt*	*NoPLA*	*High*	*regSINI*	45	45.1	0.49	0.50
High	*Noncrt*	*NoPLA*	*Md&O*	*NonPLA*	151	151.0	1.13	1.13
High	*Noncrt*	*NoPLA*	*Md&O*	*regSINI*	51	51.1	0.60	0.60
High	*Noncrt*	*w/PLA*	*Elem*	*NonPLA*	48	48.2	1.34	1.34
High	*Noncrt*	*w/PLA*	*High*	*NonPLA*	37	36.8	1.16	1.18
High	*Noncrt*	*w/PLA*	*Md&O*	*NonPLA*	7	7.4	0.90	0.98
High	*Noncrt*	*w/PLA*	*Md&O*	*regSINI*	6	6.3	0.83	0.88
High	*Noncrt*	*w/PLA*	*Md&O*	*PLA*	16	15.4	0.86	1.06
High	*Cert*	*NoPLA*	*Elem*	*NonPLA*	99	99.1	0.90	0.90
High	*Cert*	*NoPLA*	*Elem*	*regSINI*	34	34.0	0.50	0.50

Table 4-1: Span-performance Strata Cells with Balance Results from 100-Sample Simulation (Continued)

Poverty stratum	*District certainty stratum*	*PLA status*	*Span*	*PLA/SINI status*	*Assigned sample size*	*Mean*	*Std Dev*	*Root MSE*
High	*Cert*	*NoPLA*	*High*	*NonPLA*	45	45.1	0.41	0.41
High	*Cert*	*NoPLA*	*High*	*regSINI*	33	32.9	0.57	0.59
High	*Cert*	*NoPLA*	*Md&O*	*NonPLA*	34	34.0	0.32	0.32
High	*Cert*	*NoPLA*	*Md&O*	*regSINI*	24	24.0	0.49	0.49
High	*Cert*	*w/PLA*	*Elem*	*NonPLA*	114	114.1	1.02	1.02
High	*Cert*	*w/PLA*	*Elem*	*regSINI*	66	66.0	1.15	1.15
High	*Cert*	*w/PLA*	*Elem*	*PLA*	46	46.0	0.65	0.65
High	*Cert*	*w/PLA*	*High*	*NonPLA*	37	37.0	0.95	0.95
High	*Cert*	*w/PLA*	*High*	*regSINI*	29	29.0	0.70	0.70
High	*Cert*	*w/PLA*	*High*	*PLA*	127	126.7	0.76	0.81
High	*Cert*	*w/PLA*	*Md&O*	*NonPLA*	35	35.3	0.76	0.81
High	*Cert*	*w/PLA*	*Md&O*	*regSINI*	28	28.0	0.75	0.75
High	*Cert*	*w/PLA*	*Md&O*	*PLA*	46	45.9	0.70	0.71
Low	*Noncrt*	*NoPLA*	*Elem*	*NonPLA*	511	511.1	0.77	0.78
Low	*Noncrt*	*NoPLA*	*Elem*	*regSINI*	63	62.9	0.40	0.41
Low	*Noncrt*	*NoPLA*	*High*	*NonPLA*	352	351.9	0.74	0.75
Low	*Noncrt*	*NoPLA*	*High*	*regSINI*	16	16.0	0.00	0.00
Low	*Noncrt*	*NoPLA*	*Md&O*	*NonPLA*	286	286.1	0.46	0.46
Low	*Noncrt*	*NoPLA*	*Md&O*	*regSINI*	32	32.0	0.22	0.22
Low	*Noncrt*	*w/PLA*	*Elem*	*NonPLA*	34	34.1	1.67	1.68
Low	*Noncrt*	*w/PLA*	*Elem*	*regSINI*	14	14.1	1.37	1.37
Low	*Noncrt*	*w/PLA*	*Elem*	*PLA*	8	8.0	0.82	0.82
Low	*Noncrt*	*w/PLA*	*High*	*NonPLA*	20	20.1	1.01	1.01
Low	*Noncrt*	*w/PLA*	*High*	*regSINI*	4	4.1	0.63	0.64
Low	*Noncrt*	*w/PLA*	*High*	*PLA*	24	23.5	0.87	0.99
Low	*Noncrt*	*w/PLA*	*Md&O*	*NonPLA*	16	16.2	0.90	0.91
Low	*Noncrt*	*w/PLA*	*Md&O*	*regSINI*	4	3.9	0.81	0.82
Low	*Noncrt*	*w/PLA*	*Md&O*	*PLA*	20	20.0	0.88	0.88
Low	*Cert*	*NoPLA*	*Elem*	*NonPLA*	60	59.9	1.19	1.19
Low	*Cert*	*NoPLA*	*Elem*	*regSINI*	3	3.3	0.72	0.76
Low	*Cert*	*NoPLA*	*High*	*NonPLA*	36	35.7	1.12	1.16
Low	*Cert*	*NoPLA*	*High*	*regSINI*	1	0.9	0.57	0.58
Low	*Cert*	*NoPLA*	*Md&O*	*NonPLA*	28	28.0	1.29	1.29
Low	*Cert*	*NoPLA*	*Md&O*	*regSINI*	2	2.3	0.92	0.96
Low	*Cert*	*w/PLA*	*Elem*	*NonPLA*	41	41.0	1.40	1.40
Low	*Cert*	*w/PLA*	*Elem*	*regSINI*	16	15.7	1.39	1.42
Low	*Cert*	*w/PLA*	*Elem*	*PLA*	4	4.2	0.54	0.59
Low	*Cert*	*w/PLA*	*High*	*NonPLA*	24	24.3	0.91	0.96
Low	*Cert*	*w/PLA*	*High*	*regSINI*	4	3.9	0.78	0.79
Low	*Cert*	*w/PLA*	*High*	*PLA*	29	28.1	1.47	1.72
Low	*Cert*	*w/PLA*	*Md&O*	*NonPLA*	20	19.8	1.56	1.58
Low	*Cert*	*w/PLA*	*Md&O*	*regSINI*	8	8.4	1.02	1.09
Low	*Cert*	*w/PLA*	*Md&O*	*PLA*	6	6.6	1.05	1.18
Total	*Total*	*Total*	*Total*	*Total*	3,378	3378.0		

Table 4-2: Span-size Strata Cells with Balance Results from 100-Sample Simulation

Poverty stratum	*District certainty stratum*	*PLA status*	*Span*	*Size*	*Assigned sample size*	*Simulation*		
						Mean	*Std error*	*Root MSE*
High	*Noncrt*	*NoPLA*	*Elem*	*Large*	127	126.8	0.94	0.96
High	*Noncrt*	*NoPLA*	*Elem*	*Small*	232	232.1	1.08	1.09
High	*Noncrt*	*NoPLA*	*High*	*Large*	21	21.0	0.57	0.57
High	*Noncrt*	*NoPLA*	*High*	*Small*	199	199.1	0.89	0.89
High	*Noncrt*	*NoPLA*	*Md&O*	*All*	202	202.1	0.96	0.96
High	*Noncrt*	*w/PLA*	*Elem*	*Large*	17	17.1	0.99	0.99
High	*Noncrt*	*w/PLA*	*Elem*	*Small*	31	31.1	1.18	1.18
High	*Noncrt*	*w/PLA*	*High*	*Large*	37	35.8	1.16	1.18
High	*Noncrt*	*w/PLA*	*Md&O*	*All*	29	29.1	1.28	1.28
High	*Cert*	*NoPLA*	*Elem*	*Large*	86	86.0	0.39	0.39
High	*Cert*	*NoPLA*	*Elem*	*Small*	47	47.0	0.53	0.53
High	*Cert*	*NoPLA*	*High*	*Large*	55	54.9	0.61	0.62
High	*Cert*	*NoPLA*	*High*	*Small*	23	23.0	0.39	0.39
High	*Cert*	*NoPLA*	*Md&O*	*All*	58	58.0	0.40	0.40
High	*Cert*	*w/PLA*	*Elem*	*Large*	125	125.2	0.96	0.98
High	*Cert*	*w/PLA*	*Elem*	*Small*	101	100.9	1.06	1.06
High	*Cert*	*w/PLA*	*High*	*Large*	46	45.7	0.94	0.97
High	*Cert*	*w/PLA*	*High*	*Small*	147	147.0	0.95	0.95
High	*Cert*	*w/PLA*	*Md&O*	*All*	109	109.2	0.85	0.86
Low	*Noncrt*	*NoPLA*	*Elem*	*Large*	237	237.0	0.50	0.50
Low	*Noncrt*	*NoPLA*	*Elem*	*Small*	337	337.0	0.79	0.79
Low	*Noncrt*	*NoPLA*	*High*	*Large*	162	162.0	0.28	0.28
Low	*Noncrt*	*NoPLA*	*High*	*Small*	206	205.9	0.81	0.81
Low	*Noncrt*	*NoPLA*	*Md&O*	*All*	318	318.1	0.47	0.47
Low	*Noncrt*	*w/PLA*	*Elem*	*Large*	23	22.9	2.02	2.02
Low	*Noncrt*	*w/PLA*	*Elem*	*Small*	33	33.3	1.38	1.40
Low	*Noncrt*	*w/PLA*	*High*	*Large*	22	22.0	0.86	0.86
Low	*Noncrt*	*w/PLA*	*High*	*Small*	26	25.8	1.09	1.11
Low	*Noncrt*	*w/PLA*	*Md&O*	*All*	40	40.1	1.22	1.22
Low	*Cert*	*NoPLA*	*Elem*	*Large*	50	50.1	1.13	1.14
Low	*Cert*	*NoPLA*	*Elem*	*Small*	13	13.1	0.84	0.85
Low	*Cert*	*NoPLA*	*High*	*Large*	31	31.0	1.29	1.29
Low	*Cert*	*NoPLA*	*High*	*Small*	6	5.6	1.14	1.21
Low	*Cert*	*NoPLA*	*Md&O*	*All*	30	30.2	1.25	1.27
Low	*Cert*	*w/PLA*	*Elem*	*Large*	40	40.1	1.32	1.33
Low	*Cert*	*w/PLA*	*Elem*	*Small*	21	20.8	1.45	1.46
Low	*Cert*	*w/PLA*	*High*	*Large*	42	42.1	1.06	1.06
Low	*Cert*	*w/PLA*	*High*	*Small*	15	14.2	1.59	1.76
Low	*Cert*	*w/PLA*	*Md&O*	*All*	34	34.7	1.03	1.27
All	*All*	*All*	*All*	*All*	3,378	3,378.0		

Table 4-3: Simulation Empirical Probabilities for Bins Determined by Final Selection Probability

Bin center	*Bin prob minimum*	*Bin prob mean*	*Bin prob maximum*	*Bin sum of MOS*	*Frame size*	*Mean of simulation totals*	*Mean simulation sample pct*
0.25%	0.00%	0.24%	0.50%	5.0	2,091	5.3	0.25%
1.00%	0.50%	1.05%	1.50%	62.0	5,908	60.8	1.03%
2.00%	1.50%	1.97%	2.50%	125.4	6,369	127.2	2.00%
3.00%	2.50%	2.98%	3.50%	130.1	4,371	128.2	2.93%
4.00%	3.50%	3.98%	4.50%	134.7	3,386	134.7	3.98%
5.00%	4.50%	4.98%	5.50%	119.7	2,404	119.4	4.97%
6.00%	5.50%	5.98%	6.50%	104.3	1,744	106.8	6.12%
7.00%	6.50%	6.99%	7.50%	90.7	1,298	92.3	7.11%
8.00%	7.50%	7.99%	8.50%	76.3	955	76.5	8.01%
9.00%	8.50%	8.98%	9.50%	73.1	814	73.7	9.06%
10.00%	9.50%	9.99%	10.50%	66.3	664	64.2	9.67%
11.00%	10.50%	10.99%	11.50%	67.0	609	67.7	11.11%
12.00%	11.50%	12.00%	12.50%	64.9	541	63.8	11.79%
13.00%	12.50%	13.00%	13.50%	67.2	517	67.4	13.04%
14.00%	13.50%	14.02%	14.50%	66.2	472	67.4	14.27%
15.00%	14.50%	14.99%	15.50%	61.2	408	62.3	15.27%
16.00%	15.50%	16.01%	16.50%	60.5	378	62.5	16.53%
17.00%	16.50%	16.99%	17.50%	56.1	330	56.0	16.95%
18.00%	17.50%	17.99%	18.50%	57.4	319	57.4	17.99%
19.00%	18.50%	19.00%	19.50%	50.5	266	51.2	19.25%
20.00%	19.52%	19.99%	20.50%	45.8	229	45.4	19.83%
21.00%	20.50%	21.00%	21.50%	47.7	227	48.5	21.35%
22.00%	21.51%	22.01%	22.50%	49.1	223	49.3	22.11%
23.00%	22.50%	23.00%	23.50%	47.6	207	47.5	22.92%
24.00%	23.50%	23.99%	24.49%	48.2	201	48.5	24.14%
25.00%	24.51%	24.98%	25.50%	45.0	180	43.8	24.31%
26.00%	25.50%	25.99%	26.50%	42.1	162	41.0	25.33%
27.00%	26.50%	26.97%	27.50%	36.7	136	36.1	26.54%
28.00%	27.50%	27.99%	28.49%	43.4	155	43.9	28.34%
29.00%	28.50%	28.98%	29.50%	39.7	137	39.6	28.90%
30.00%	29.52%	29.97%	30.49%	38.1	127	38.5	30.31%
31.00%	30.50%	31.01%	31.50%	45.0	145	46.2	31.83%
32.00%	31.51%	31.95%	32.48%	31.0	97	31.6	32.57%
33.00%	32.52%	33.00%	33.49%	40.9	124	39.6	31.94%
34.00%	33.50%	33.96%	34.49%	25.1	74	24.0	32.41%
35.00%	34.54%	34.95%	35.48%	21.0	60	22.0	36.68%
36.00%	35.51%	35.99%	36.50%	28.4	79	29.3	37.06%
37.00%	36.50%	37.00%	37.50%	23.3	63	23.1	36.60%
38.00%	37.52%	38.06%	38.50%	23.6	62	23.0	37.05%
39.00%	38.50%	39.01%	39.46%	21.8	56	20.9	37.36%
40.00%	39.51%	40.06%	40.49%	20.0	50	20.3	40.62%
41.00%	40.51%	41.00%	41.50%	28.7	70	27.7	39.57%
42.00%	41.52%	42.02%	42.49%	32.8	78	31.4	40.24%
43.00%	42.53%	43.06%	43.49%	37.9	88	37.8	43.00%

Table 4-3: Simulation Empirical Probabilities for Bins Determined by Final Selection Probability (Continued)

Bin center	*Bin prob minimum*	*Bin prob mean*	*Bin prob maximum*	*Bin sum of MOS*	*Frame size*	*Mean of simulation totals*	*Mean simulation sample pct*
44.00%	43.51%	44.00%	44.46%	31.7	72	32.0	44.44%
45.00%	44.50%	45.03%	45.50%	36.5	81	36.2	44.69%
46.00%	45.50%	45.97%	46.49%	32.2	70	32.7	46.70%
47.00%	46.50%	46.96%	47.49%	37.1	79	36.6	46.27%
48.00%	47.51%	47.99%	48.49%	33.6	70	33.4	47.74%
49.00%	48.54%	49.03%	49.49%	23.0	47	22.8	48.60%
50.00%	49.60%	50.04%	50.48%	27.0	54	27.6	51.15%
51.00%	50.52%	51.03%	51.48%	19.4	38	19.1	50.29%
52.00%	51.51%	51.89%	52.49%	17.6	34	17.5	51.50%
53.00%	52.53%	53.03%	53.49%	24.9	47	25.8	54.81%
54.00%	53.51%	54.04%	54.49%	24.9	46	25.0	54.33%
55.00%	54.52%	54.90%	55.48%	24.7	45	24.2	53.87%
56.00%	55.50%	55.97%	56.49%	22.4	40	22.6	56.50%
57.00%	56.50%	56.96%	57.48%	37.0	65	38.5	59.28%
58.00%	57.51%	58.02%	58.48%	28.4	49	28.5	58.16%
59.00%	58.54%	59.05%	59.49%	23.6	40	24.2	60.53%
60.00%	59.50%	59.94%	60.47%	35.4	59	34.8	58.92%
61.00%	60.54%	61.00%	61.49%	34.2	56	33.9	60.52%
62.00%	61.54%	61.97%	62.50%	33.5	54	33.0	61.04%
63.00%	62.51%	63.01%	63.50%	43.5	69	43.0	62.28%
64.00%	63.53%	64.06%	64.47%	26.9	42	26.2	62.29%
65.00%	64.50%	64.95%	65.49%	33.8	52	35.0	67.38%
66.00%	65.52%	66.00%	66.46%	29.0	44	29.2	66.27%
67.00%	66.58%	67.00%	67.47%	26.8	40	25.8	64.53%
68.00%	67.53%	68.02%	68.45%	15.6	23	15.1	65.48%
69.00%	68.51%	69.02%	69.45%	17.9	26	18.0	69.38%
70.00%	69.52%	69.97%	70.44%	16.1	23	16.2	70.26%
71.00%	70.51%	70.93%	71.44%	16.3	23	15.8	68.65%
72.00%	71.60%	72.00%	72.50%	15.8	22	15.5	70.45%
73.00%	72.73%	73.11%	73.41%	8.0	11	8.0	73.09%
74.00%	73.67%	73.97%	74.38%	7.4	10	7.3	73.00%
75.00%	74.53%	75.11%	75.50%	9.0	12	8.8	73.17%
76.00%	75.51%	76.06%	76.48%	12.9	17	12.7	74.94%
77.00%	76.57%	77.00%	77.49%	8.5	11	8.2	74.82%
78.00%	77.52%	77.91%	78.36%	6.2	8	5.9	74.00%
79.00%	78.53%	78.86%	79.23%	4.7	6	5.0	83.67%
80.00%	79.55%	80.07%	80.48%	8.8	11	8.7	78.91%
81.00%	80.64%	81.06%	81.38%	6.5	8	6.6	83.00%
82.00%	81.55%	82.07%	82.46%	6.6	8	6.5	81.00%
83.00%	82.65%	83.00%	83.25%	4.2	5	4.3	85.40%
84.00%	83.58%	84.13%	84.48%	6.7	8	6.9	85.88%
85.00%	84.53%	84.80%	85.16%	5.1	6	5.0	82.83%
86.00%	85.72%	86.08%	86.37%	5.2	6	5.0	82.67%
87.00%	86.60%	87.03%	87.39%	10.4	12	10.7	88.92%

Table 4-3: Simulation Empirical Probabilities for Bins Determined by Final Selection Probability (Continued)

Bin center	*Bin prob minimum*	*Bin prob mean*	*Bin prob maximum*	*Bin sum of MOS*	*Frame size*	*Mean of simulation totals*	*Mean simulation sample pct*
88.00%	87.60%	87.98%	88.44%	6.2	7	6.2	88.86%
89.00%	88.50%	89.13%	89.37%	6.2	7	6.4	91.71%
90.00%	89.77%	90.16%	90.49%	5.4	6	5.3	89.00%
91.00%	90.50%	91.00%	91.48%	11.8	13	11.6	89.23%
92.00%	91.81%	92.10%	92.43%	4.6	5	4.4	88.00%
93.00%	92.64%	93.15%	93.38%	6.5	7	6.5	92.29%
94.00%	93.56%	93.85%	94.26%	5.6	6	5.8	96.83%
95.00%	94.75%	95.08%	95.46%	4.8	5	4.7	94.20%
96.00%	95.51%	96.06%	96.40%	6.7	7	6.6	93.71%
97.00%	96.52%	96.63%	96.78%	5.8	6	5.7	95.50%
98.00%	97.59%	98.01%	98.40%	6.9	7	6.8	96.86%
99.00%	98.54%	98.83%	99.07%	5.9	6	5.8	96.00%
99.75%	99.59%	99.59%	99.59%	1.0	1	1.0	100.00%

The results of the simulation study appear to be quite favorable. The variability in the final sample sizes for the span-performance and span-size strata are quite limited. The true inclusion probabilities appear to be very close to the assigned probabilities, at least in aggregate within bins defined by 1% intervals for the assigned probabilities. We don't know if there are particular frame schools whose true inclusion probabilities deviate from the assigned probabilities, though we can say the number of these and level of deviation are likely to be quite limited, given the aggregate results.

References

Bryant, E. C., Hartley, H. O., Jessen, R. T. (1960). Design and estimation in two-way stratification. *Journal of the American Statistical Association 55*, 105-124.

Causey, B. D., Cox, L. H., and Ernst, L. R. (1985). Applications of transportation theory to statistical problems. *Journal of the American Statistical Association 80*, 903-909.

Deville, J.-C., and Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika 91*, 4, 893-912.

Deville, J.-C., and Tillé, Y. (2005). Variance approximation under balanced sampling. *Journal of Statistical Planning and Inference* 128, 569-591.

Falorsi, P. D., and Righi, P. (2008). A balanced sampling approach for multi-way stratification designs for small area estimation. *Survey Methodology 34, 2*, 223-234.

Goodman, R., and Kish, L. (1950). Controlled selection—A technique in probability sampling. *Journal of the American Statistical Association 45*, 350-372.

Groves, R. M., and Hess, I. (1975). An algorithm for controlled selection, in *Probability Sampling of Hospitals and Patients, 2nd edition* (Eds. I. Hess, D. C. Riedel, T. B. Fitzpatrick). Ann Arbor, MI: Health Administration Press.

Jessen, R. T. (1970). Probability sampling with marginal constraints. *Journal of the American Statistical Association 65*, 776-796.

Kish, L. (1992). Weighting for unequal Pi. *Journal of Official Statistics 8, 2*, 183-200.

Sitter, R. R., and Skinner, C. J. (1994). Multi-way stratification by linear programming. *Survey Methodology 20*, 65-73.