GUIDANCE ON SELECTING A SAMPLE FOR A SCHOOL-BASED ORAL HEALTH SURVEY
MAY 2013, UPDATED JUNE 2015 AND JULY 2017

Is your state, territory or local health agency planning to conduct a school-based oral health survey?

If yes, then you undoubtedly have questions about how to select an appropriate sample. This topic is important because proper sampling design and methods are crucial for valid population estimates and statistical assessments of precision of estimates. The purpose of this document is to give you a basic framework for how a sample of schools is selected. Although this document is geared towards states and territories, the techniques are appropriate for other jurisdictions such as counties.

Because no one method is appropriate for all states/territories we encourage you to read this document then contact ASTDD for additional guidance on selecting a sample for your state/territory.

This document is limited to a discussion of sampling. For additional information on how to conduct and use data from a school-based oral health survey, please refer to the Basic Screening Survey (BSS) tools developed by the Association of State and Territorial Dental Directors (ASTDD). These tools are available at the following website: www.astdd.org/basic-screening-survey-tool/.

Do you want to submit the data to the National Oral Health Surveillance System (NOHSS)?

NOHSS (www.cdc.gov/oralhealthdata) is a collaborative effort between CDC's Division of Oral Health and ASTDD. NOHSS is designed to monitor the burden of oral disease, use of the oral health care delivery system, and the status of community water fluoridation on both a national and state level. NOHSS is designed to track oral health surveillance indicators based on data sources and surveillance capacity available to most states. The Council of State and Territorial Epidemiologists (CSTE) and the Chronic Disease Directors (CDD) were instrumental in developing the framework for chronic disease surveillance indicators, including the oral health indicators in NOHSS.

If you follow the guidance provided in this document your oral health data will meet the specifications for inclusion in NOHSS.

Only data that meet the following specifications are included in the NOHSS data system:

- The data are from a ***statewide probability sample*** of elementary schools.
- If a complex sampling scheme is used, the data must be weighted for the sampling scheme.
- ASTDD strongly suggests that, at minimum, 3rd grade children be screened. Grades K-2 as well as Head Start may also be screened and are included on the NOHSS website.

Why not have just one sampling plan for all states?

There are limitations to creating one single sample design for all states. States differ in size, geography, political boundaries, population distribution, population demographics and in the make-up and arrangement of the schools to be sampled. In addition to the basic BSS oral health indicators for the state population, state oral health programs may have additional indicators or population subgroups of specific interest. Resources may also vary, affecting the size and extent of BSS surveys, which in turn can affect the sample design. These differences may require unique sample design features, limiting the degree to which a single predefined sampling plan fits all situations. Given these considerations, some general guidelines for BSS survey sampling follow.

What basic sampling guidelines should I follow?

The following guidelines are intended to aid states in designing sampling plans for oral health surveys using the BSS methodology. These guidelines focus on choosing target populations and employing appropriate stratification and cluster sample selection techniques. Schools represent natural clusters for sample selection. Effective stratification, based on geographic area or the proportion of students participating in the National School Lunch Program (NSLP) for example, can improve the statistical efficiency of sampling. Such stratification can be implemented, along with appropriate cluster selection, in the sampling schemes described below. Using common sampling techniques in combination with BSS methodology can improve the efficiency of state oral health needs assessment surveys and improve precision and reduce bias in estimates and tracking of trends. Standardizing methods of sampling can also increase comparability of BSS findings between states.

What sampling designs are appropriate for a school-based oral health survey?

There are many different types of sampling designs but some, such as simple random sampling, are not appropriate for a school-based oral health survey. Following are types of sampling designs that are appropriate.

When implementing survey sampling methodology, follow these basic guidelines:

- ***At a minimum, school surveys should include 3rd graders in public on-site schools. States can determine if they want to include other types of schools such as private schools or schools administered by the Bureau of Indian Education.***
- ***Sample designs for school oral health surveys should ensure good representation on socioeconomic status (SES) through stratification (preferably implicit) on NSLP participation at the school level.***
- ***Replacement schools should be selected for schools that refuse to participate through a random probability selection process from schools in the same stratum or sampling interval.***
- ***Selection probabilities and response rates should be tracked for use in calculating sampling weights for data analysis.***

Each of these designs employ cluster sampling for efficiencies (sampling schools vs. random sample of children).

- **Stratified random sampling**: Separate the sampling frame into categories (e.g., counties, health regions, urban/rural) and randomly select from each category.
  - Proportionate – select a consistent fraction from each group to sample proportionally. The numbers selected into the sample will be proportional to the numbers in the population groups (strata).
  - Disproportionate – select at different rates from the groups, e.g., only 10% of the population is in a particular category, but you are particularly interested in that category (e.g., rural counties) so you sample at a higher rate to get better estimates. Oversampling would be used for strata in which proportionate sampling might not result in selection of a large enough number to obtain sufficiently precise population estimates.
- **Systematic sampling**: Sampling from a list using a selection interval and random start. For example, sampling every 6th element from a list starting with a random number between 1 and 6.
  - **NOTE:** Ordering the list by group (strata) and then using a systematic sample makes this method equivalent to (or better than) a proportionate stratified sample because this ***implicit stratification*** can be used with continuous variables (e.g., ordering schools by NSLP percentage) to provide finer stratification and can be combined with probability proportional to size (PPS) sampling.
- **Probability proportional to size (PPS) sampling**: Sampling of clusters (schools) proportional to their size. Proportional to size sampling is often employed in different stages of a multistage sample. For example, in a survey of 3rd graders, school districts could be sampled according to total 3rd grade enrollment size, and then schools within selected districts could be sampled proportional to school 3rd grade enrollment. All children in targeted grades in selected schools can be included in the sample, but if a set number (e.g., a typical class size of 25) is sampled within selected schools, the sample can become self-weighting, meaning the higher probability of selection for larger districts and schools is offset by the lower probability of selecting children within the large schools so all children in the survey population

have about the same probability of selection. This self-weighting is for probability of selection, and doesn't account for non-response (which will vary by school). Analysis weights may be needed to account for non-response. PPS sampling can be advantageous because it ensures sufficient representation of the larger clusters (schools). This may increase the efficiency of the survey logistically, as larger clusters (schools) are often more geographically concentrated in urban areas, thus decreasing survey travel time and costs.

What steps should I follow when selecting a sample?

Step 1. Determine which indicators will be collected. Refer to the BSS manual for a list of the recommended indicators (www.astdd.org/basic-screening-survey-tool/).

Step 2. Identify your target population and any sub-populations of interest. For example, do you want oral health estimates for 3rd grade children for the state as a whole or do you want oral health estimates for 3rd grade children by region or by county? Other sub-populations of interest could include a particular geographic area, racial/ethnic minorities, low-income or rural children.

Step 3. Define the survey population to include in your sampling frame considering the practical limitations for accessing the entire target population. For example, you may decide to restrict your sampling frame to on-site public schools with 20 or more children in 3rd grade. Restricting your sampling frame helps to assure that resources aren't wasted by going to schools with a very small number of participants.

Step 4. Determine your sample size based on the population level estimates desired (e.g. state, region, health district, county) and the level of statistical precision desired. Following are some general approximate guidelines to sample size determination.

a. BSS indicators are proportions or percent estimates (e.g. percent of children with untreated tooth decay)
   i. The simple random sampling formula for proportions is: $v(p_0) = (1-f)\ p_0q_0/n-1$
   ii. The most conservative calculation is to use estimated p_0 = 50%. This results in sample sizes associated with the levels of desired precision outlined in Box 1.

b. Multiply the sample size by the estimated design effect, which reflects the effects of complex sample design vs. simple random sampling on variance/precision estimation (often about 2).

c. **NOTE:** The same sample size calculations are appropriate for the population level of interest. If you want regional level estimates and there are 5 regions in the state, then you should multiply the sample size by 5 (n X 5). If county level estimates are desired and there are 35 counties in the state, then you should multiply the sample size by 35 (n X 35). For example, if you decided that you wanted estimates with a 95% confidence interval of +/- 10% for 35 counties, you would need a sample size of 97 from each of the 35 counties, for a total sample size of 35*97=3,395.

d. **NOTE:** Available resources often drive sample size determination to a greater extent than statistical precision considerations. Available funds, time and number of trained screeners, which affect the number of schools you are able to screen, may be the limiting factors in determining sample size.

Box 1

For 95% CI = +/- 10% n = 97
For 95% CI = +/- 5% n = 384
For 95% CI = +/- 3% n = 1,066

Step 5. Prepare your sampling frame. Obtain an electronic list of schools from your state department of education. At a minimum, the file should include school name, school ID code, district name, district ID code, enrollment by grade, total enrollment, county or region, and number or percent of children participating in NSLP. Other useful information includes enrollment by race/ethnicity which allows you to determine if the sample is representative of the state in terms of race/ethnicity. To make contacting the school easier, it is also useful to have the school's address, phone number and email address.

NOTE: The National School Lunch Program is a federally assisted meal program operating in public and nonprofit private schools and residential child care institutions. Each state may use a different name for this program. For example, West Virginia refers to this as "percent needy". The federal Healthy, Hunger-Free Kids Act of 2010 established a community eligibility provision (CEP) within the NSLP. This is a reimbursement option for eligible educational agencies and schools that allows them to offer free school meals to all children in high poverty schools without collecting household applications. In states where schools are using the CEP option, you may need to obtain information on both NSLP and CEP. For additional information on CEP refer to http://www.fns.usda.gov/school-meals/community-eligibility-provision.

Step 6. Stratify the sampling frame. This helps to ensure representation of population subgroups of interest or importance and almost always improves precision of overall survey estimates.

a. Common stratification variables include:
    i. Geographic factors such as county, region or health district
    ii. Urban/rural status
    iii. NSLP status of school
b. Most states use multiple levels of stratification. For example, a state may stratify by region, by urban/rural within region, and by NSLP percent within urban/rural.
c. Stratification can be:
    i. Explicit – sampling within each stratum
    ii. Implicit – with systematic sampling from a list sorted by stratification variables
d. **NOTE:** ASTDD recommends that all states, at a minimum, use stratification (preferably implicit) by the percent of children that participate in NSLP.

Step 7. Select the sample using one of the following general methods.

a. Probability proportional to size (PPS) sampling of schools. With PPS sampling, larger schools have a higher probability of being selected. For a self-weighting analysis, screen a consistent number of children (e.g. 1 or 2 classrooms) per school. PPS sampling with a consistent number of children screened at each school can result in an efficient scheduling of screeners and ensures proportionate representation of children from different sized clusters (i.e., large clusters are not under-represented).
b. Probability sampling of schools without regard to school size (non-PPS). With a non-PPS sampling of schools, each school, regardless of size has an equal probability of being selected. For a self weighting analysis, all children in the target grade should be screened. **NOTE:** The sample should be self weighting within the strata but you will need to include an adjustment factor to account for differences in stratum population sizes.

What are some examples of the sampling process?

Following are two examples to help you visualize the sampling process. The first example uses PPS sampling while the second example uses a non-PPS sampling design. Both examples generate a sample that is implicitly stratified by region, urban/rural status and percent of children participating in NSLP. If systematic sampling will be part of your PPS sampling scheme, then a cumulative running total of school 3rd grade child enrollment should be included starting with the first school and adding through the entire sampling frame.

NOTE: The following examples use Excel for setting up the sampling frame and selecting the sample. Automated procedures, such as SAS SurveySelect may present difficulties for sampling techniques such as systematic selection from an ordered list to achieve implicit stratification and for selection of replacements for refusing schools, which is very important in assuring that the final surveyed sample is representative of the population.

Example #1: Systematic PPS sampling with implicit stratification by region, urban/rural status and F/R lunch

This example describes a PPS sampling strategy. In this example, larger schools have a higher probability of selection. Based on available resources, the decision was made to include 70 schools in the "Utopia" oral health survey of 3rd grade children. The following sampling steps were employed:

- The sampling frame list was sorted by region then by urban/rural status within each region
- Schools were then sorted by percent of children participating in NSLP within urban/rural school categories.

Calculations used for systematic PPS sampling are as follows:

- Sampling interval for sampling = (total 3rd grade enrollment) / (# of schools to be screened)
  - 53,320 / 70 = 761.7
- Random start = random number between 0 and interval (761.7) = 148.0
  - This is the first school selection number
  - There are a variety of methods for selecting a random number including, but not limited to, Excel and www.random.org
- Select the school with the 148th child. Add the sampling interval (761.7) to 148 to get the next school (909.7). Continue adding the sampling interval repeatedly until all 70 school selections are made.

  148.0, 909.7, 1671.4, 2433.1, 3194.8, 3956.5, 4718.2, …
- These numbers are matched to the cumulative enrollment numbers in the sampling list. The schools with enrollment intervals containing the sample selection numbers are selected into the sample. The sampling frame list and the selected schools are shown in Table 1.

Example #2: Systematic sampling with implicit stratification by region, urban/rural status and F/R lunch

This example describes a non-PPS sampling strategy. In this example, all schools (regardless of size) have the same probability of selection. Based on available resources, the decision was made to include 70 schools in the "Utopia" oral health survey of 3rd grade children. The following sampling steps were employed:

- The sampling frame list was sorted by region then by urban/rural status within each region
- Schools were then sorted by percent of children participating in NSLP within urban/rural school categories.

Calculations used for systematic sampling are as follows:

- Sampling interval for sampling = (number of schools in sampling frame) / (# of schools to be screened)
  - 700 / 70 = 10.0
- Random start = random number between 1 and interval (10) = 6
  - This is the first school selection number
  - There are a variety of methods for selecting a random number including, but not limited to, Excel and www.random.org
- Select the 6th school. Add the sampling interval (10.0) to 6 to get the next school (16.0). Continue adding the sampling interval repeatedly until all 70 school selections are made.

  6.0, 16.0, 26.0, 36.0, 46.0, 56.0, 66.0, …
- These numbers are matched to the sequential number of schools in the sampling list to identify the schools selected into the sample. The sampling frame list and the selected schools are shown in Table 2.
- **NOTE:** In this example, dividing the number of schools by the number to screen produced a whole number. Please contact ASTDD if you need more information on how to select a non-PPS sample when the sampling interval includes a decimal.

What should I do if a school refuses to participate?

An important part of the sample design is the replacement of refusing clusters (schools). Refusals should be replaced with similar probability methods as original selections. Selection of a replacement should be from the same sampling interval as the refusing school so that the sampling interval is represented.

- If systematic PPS sampling was used in the original sample, a PPS method of replacement selection should be used. Table 3 shows the method of replacement used in the Utopia PPS sampling example.
  - The list of the final selections is created that includes enrollment size and start of the sampling interval on the sampling frame list.
  - The enrollment size of the refusing school is subtracted from the sampling interval size.
  - A random number between 0-1 is generated and applied to the remaining sample interval, using the sample interval start to determine the position of the replacement selection in the interval. The following website can be used for selecting a random number between 0-1: http://www.random.org/decimal-fractions/
  - The original sample frame list is then viewed to see where this replacement number falls, to determine the replacement school.
  - **NOTE:** If your replacement number calculated is equal to or greater than the refusal school interval, you have to add the enrollment number for the refusing school for your replacement selection number (e.g. the second replacement in Table 3 - 133 is added for a final replacement number of 3531.2). This adjusts for the fact that the refusing school is no longer in the adjusted interval.
  - Table 3 shows replacement numbers and calculations for two refusing schools that were originally selected using PPS.
- If systematic non-PPS sampling was used in the original sample, a non-PPS method of replacement selection should be used. Table 4 shows the method of replacement used in the Utopia non-PPS sampling example.
  - Determine the range of the sampling interval for the refusing school.
  - Select a random number between the lowest and highest school number in the interval. If you get a number matching the refusing school, just generate another number.
  - The school with that number is the replacement.

Will I need specialized software to analyze the data?

Yes, you will need to use statistical software designed to address the statistical ramifications of complex probability sample designs, specifically stratification and cluster sampling. Appropriate software packages include, but are not limited to, SUDAAN and survey analysis procedures in SAS, Stata, SPSS, Epi Info and R.

Where can I get additional help?

ASTDD can help you with the sample selection process. Please contact us if you have any questions.

Association of State & Territorial Dental Directors
Kathy Phipps, Data and Surveillance Coordinator
Phone: 805-776-3393, Email: kathyphipps1234@gmail.com

Acknowledgements

Supported by Cooperative Agreement NU5U8DP004919 from the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of CDC. ASTDD would like to thank Michael Manz, Kathy Phipps and Laurie Barker for their assistance in developing and reviewing this guidance.

References and additional resources

- Kalton G (1983). Introduction to survey sampling, Sage, Beverly Hills, CA.
- Kish L (1965). Survey Sampling, John Wiley & Sons, New York.
- Heeringa SG, West BT, Berglund PA (2010). Applied survey data analysis, Chapman and Hall, London.

Table 1: Systematic PPS sampling with implicit stratification by region, urban/rural and NSLP participation

Region	Urban/ Rural	School Name	National School Lunch Program Percent	3rd Grade Enrollment	Cumulative 3rd Grade Enrollment	Selected School
1	Rural	KEOWEE	37.1%	38	38	
1	Rural	WALHALLA	39.8%	92	130	
1	**Rural**	**RAVENEL**	**52.3%**	**91**	**221**	**148.0**
1	Rural	LAKEVIEW	52.4%	94	315	
1	Rural	NINETY SIX	52.7%	130	445	
1	Rural	NORTHSIDE	52.8%	95	540	
1	Rural	MCCORMICK	53.9%	56	596	
1	Rural	FAIR-OAK	55.7%	117	713	
1	Rural	HICKORY TAVERN	56.8%	61	774	
1	Rural	HOLLYWOOD	57.2%	66	840	
1	Rural	CHEROKEE TRAIL	57.8%	53	893	
1	**Rural**	**PINECREST**	**60.0%**	**100**	**993**	**909.7**
1	Rural	MERRYWOOD	60.1%	90	1,083	
1	Rural	DIAMOND HILL	60.3%	38	1,121	
1	Rural	SPRINGFIELD	60.9%	89	1,210	
1	Rural	TAMASSEE-SALEM	61.7%	41	1,251	
1	Rural	WESTMINSTER	67.2%	62	1,313	
1	Rural	HODGES	67.5%	36	1,349	
1	Rural	LAURENS	67.7%	92	1,441	
1	Rural	GRAY COURT OWINGS	68.7%	58	1,499	
1	Rural	E B MORSE	68.7%	95	1,594	
1	**Rural**	**CLINTON**	**69.0%**	**87**	**1,681**	**1,671.4**
1	Rural	WESTWOOD	69.8%	120	1,801	
1	Rural	ORCHARD PARK	71.5%	61	1,862	
1	Rural	WARE SHOALS PRIMARY	72.0%	55	1,917	
1	Rural	JOANNA-WOODSON	72.2%	50	1,967	
1	Rural	JAMES M BROWN	73.9%	98	2,065	
1	Rural	OAKLAND	74.8%	78	2,143	
1	Rural	BLUE RIDGE ELEMENTARY	76.6%	90	2,233	
1	Rural	WATERLOO	77.5%	37	2,270	
1	Rural	EASTSIDE	78.9%	67	2,337	
1	**Rural**	**WOODFIELDS**	**80.8%**	**97**	**2,434**	**2,433.1**
1	Rural	SALUDA	81.0%	107	2,541	
1	Rural	MATHEWS	83.3%	84	2,625	
1	Rural	JOHN C CALHOUN	89.9%	36	2,661	
1	Rural	FORD	92.8%	81	2,742	
1	Urban	MIDWAY SCHL	16.7%	142	2,884	
1	Urban	WREN	26.5%	100	2,984	
1	Urban	WRIGHT	28.6%	28	3,012	
1	Urban	POWDERSVILLE	31.2%	173	3,185	
1	**Urban**	**CONCORD**	**31.5%**	**133**	**3,318**	**3,194.8**
1	Urban	HUNT MEADOWS	39.8%	75	3,393	
1	Urban	MT LEBANON	41.9%	55	3,448	
1	Urban	MERRIWETHER	48.4%	120	3,568	
1	Urban	SPEARMAN	51.3%	60	3,628	
1	Urban	LA FRANCE	51.3%	52	3,680	
1	Urban	BELTON	52.2%	160	3,840	
1	Urban	STARR	53.2%	57	3,897	
1	**Urban**	**CENTERVILLE**	**55.5%**	**117**	**4,014**	**3,956.5**
1	Urban	WEST PELZER	55.6%	69	4,083	
1	Urban	HONEA PATH	55.6%	97	4,180	
1	Urban	CEDAR GROVE	55.9%	90	4,270	
1	Urban	PALMETTO	61.9%	90	4,360	
1	Urban	TOWNVILLE	63.5%	36	4,396	
1	Urban	W E PARKER	64.9%	88	4,484	
1	Urban	MCLEES	65.1%	118	4,602	
1	Urban	IVA	67.2%	70	4,672	
1	**Urban**	**CALHOUN ACADEMY**	**67.6%**	**120**	**4,792**	**4,718.2**
1	Urban	WHITEHALL	69.9%	75	4,867	
1	Urban	NEW PROSPECT	72.3%	66	4,933	
1	Urban	JOHNSTON	73.5%	49	4,982	
1	Urban	HOMELAND PARK	74.7%	52	5,034	
1	Urban	PENDLETON	74.7%	52	5,086	
1	Urban	FLAT ROCK .	76.5%	67	5,153	
1	Urban	NEVITT FOREST SCHOOL	80.7%	66	5,219	
1	Urban	DOUGLAS	82.4%	48	5,267	
1	Urban	VARENNES ACADEMY	90.8%	65	5,332	
2	Rural	SPARTANBURG	43.2%	45	5,377	
2	Rural	LOCKHART	58.8%	23	5,400	
2	Rural	BUFFALO	70.9%	103	5,503	

* In PPS sampling, the number of children in the sampling interval is the same for each interval; in this example 761.7. This number will be used to calculate the weight factor.

Table 2: Systematic sampling (non-PPS) with implicit stratification by region, urban/rural status and NSLP participation

Region	Urban/ Rural	School Name	NSLP Percent	3rd Grade Enrollment	Sequential School Number	Sampling Interval #	# of Children in Sampling Interval*	Selected School
1	Rural	KEOWEE	37.1%	38	1	1		
1	Rural	WALHALLA	39.8%	92	2	1		
1	Rural	RAVENEL	52.3%	91	3	1		
1	Rural	LAKEVIEW	52.4%	94	4	1		
1	Rural	NINETY SIX	52.7%	130	5	1		
1	**Rural**	**NORTHSIDE**	**52.8%**	**95**	**6**	**1**		**6**
1	Rural	MCCORMICK	53.9%	56	7	1		
1	Rural	FAIR-OAK	55.7%	117	8	1		
1	Rural	HICKORY TAVERN	56.8%	61	9	1		
1	Rural	HOLLYWOOD	57.2%	66	10	1	840	
1	Rural	CHEROKEE TRAIL	57.8%	53	11	2		
1	Rural	PINECREST	60.0%	100	12	2		
1	Rural	MERRYWOOD	60.1%	90	13	2		
1	Rural	DIAMOND HILL	60.3%	38	14	2		
1	Rural	SPRINGFIELD	60.9%	89	15	2		
1	**Rural**	**TAMASSEE-SALEM**	**61.7%**	**41**	**16**	**2**		**16**
1	Rural	WESTMINSTER	67.2%	62	17	2		
1	Rural	HODGES	67.5%	36	18	2		
1	Rural	LAURENS	67.7%	92	19	2		
1	Rural	GRAY COURT OWINGS	68.7%	58	20	2	659	
1	Rural	E B MORSE	68.7%	95	21	3		
1	Rural	CLINTON	69.0%	87	22	3		
1	Rural	WESTWOOD	69.8%	120	23	3		
1	Rural	ORCHARD PARK	71.5%	61	24	3		
1	Rural	WARE SHOALS	72.0%	55	25	3		
1	**Rural**	**JOANNA-WOODSON**	**72.2%**	**50**	**26**	**3**		**26**
1	Rural	JAMES M BROWN	73.9%	98	27	3		
1	Rural	OAKLAND	74.8%	78	28	3		
1	Rural	BLUE RIDGE	76.6%	90	29	3		
1	Rural	WATERLOO	77.5%	37	30	3	771	
1	Rural	EASTSIDE	78.9%	67	31	4		
1	Rural	WOODFIELDS	80.8%	97	32	4		
1	Rural	SALUDA	81.0%	107	33	4		
1	Rural	MATHEWS	83.3%	84	34	4		
1	Rural	JOHN C CALHOUN	89.9%	36	35	4		
1	**Rural**	**FORD**	**92.8%**	**81**	**36**	**4**		**36**
1	Urban	MIDWAY	16.7%	142	37	4		
1	Urban	WREN	26.5%	100	38	4		
1	Urban	WRIGHT	28.6%	28	39	4		
1	Urban	POWDERSVILLE	31.2%	173	40	4	915	
1	Urban	CONCORD	31.5%	133	41	5		
1	Urban	HUNT MEADOWS	39.8%	75	42	5		
1	Urban	MT LEBANON	41.9%	55	43	5		
1	Urban	MERRIWETHER	48.4%	120	44	5		
1	Urban	SPEARMAN	51.3%	60	45	5		
1	**Urban**	**LA FRANCE**	**51.3%**	**52**	**46**	**5**		**46**
1	Urban	BELTON	52.2%	160	47	5		
1	Urban	STARR	53.2%	57	48	5		
1	Urban	CENTERVILLE	55.5%	117	49	5		
1	Urban	WEST PELZER	55.6%	69	50	5	898	
1	Urban	HONEA PATH	55.6%	97	51	6		
1	Urban	CEDAR GROVE	55.9%	90	52	6		
1	Urban	PALMETTO	61.9%	90	53	6		
1	Urban	TOWNVILLE	63.5%	36	54	6		
1	Urban	W E PARKER	64.9%	88	55	6		
1	**Urban**	**MCLEES**	**65.1%**	**118**	**56**	**6**		**56**
1	Urban	IVA	67.2%	70	57	6		
1	Urban	CALHOUN ACADEMY	67.6%	120	58	6		
1	Urban	WHITEHALL	69.9%	75	59	6		
1	Urban	NEW PROSPECT	72.3%	66	60	6	850	
1	Urban	JOHNSTON	73.5%	49	61	7		
1	Urban	HOMELAND PARK	74.7%	52	62	7		
1	Urban	PENDLETON	74.7%	52	63	7		
1	Urban	FLAT ROCK	76.5%	67	64	7		
1	Urban	NEVITT FOREST	80.7%	66	65	7		
1	**Urban**	**DOUGLAS**	**82.4%**	**48**	**66**	**7**		**66**
1	Urban	VARENNES ACADEMY	90.8%	65	67	7		
2	Rural	SPARTANBURG	43.2%	45	68	7		
2	Rural	LOCKHART	58.8%	23	69	7		
2	Rural	BUFFALO	70.9%	103	70	7	570	

* The number of children in the sampling interval will be used to calculate the weight factor.

Table 3: Example of refusal replacement for systematic PPS sample selection

		A		B	
School Name	NSLP%	3rd Enroll.	Cumulative 3rd Grade Enroll.	Original Selection Number	Status
KEOWEE	37.1%	38	38		
WALHALLA	39.8%	92	130		
RAVENEL	**52.3%**	**91**	**221**	**148.0**	
LAKEVIEW	52.4%	94	315		
NINETY SIX	52.7%	130	445		
NORTHSIDE	52.8%	95	540		
MCCORMICK	53.9%	56	596		
FAIR-OAK	55.7%	117	713		
HICKORY TAVERN	56.8%	61	774		
HOLLYWOOD	57.2%	66	840		
CHEROKEE TRAIL	57.8%	53	893	**874.2**	**Replacement**
PINECREST	**60.0%**	**100**	**993**	**909.7**	**Refused**
MERRYWOOD	60.1%	90	1,083		
DIAMOND HILL	60.3%	38	1,121		
SPRINGFIELD	60.9%	89	1,210		
TAMASSEE-SALEM	61.7%	41	1,251		
WESTMINSTER	67.2%	62	1,313		
HODGES	67.5%	36	1,349		
LAURENS	67.7%	92	1,441		
GRAY COURT OWINGS	68.7%	58	1,499		
E B MORSE	68.7%	95	1,594		
CLINTON	**69.0%**	**87**	**1,681**	**1,671.4**	
WESTWOOD	69.8%	120	1,801		
ORCHARD PARK	71.5%	61	1,862		
WARE SHOALS	72.0%	55	1,917		
JOANNA-WOODSON	72.2%	50	1,967		
JAMES M BROWN	73.9%	98	2,065		
OAKLAND	74.8%	78	2,143		
BLUE RIDGE	76.6%	90	2,233		
WATERLOO	77.5%	37	2,270		
EASTSIDE	78.9%	67	2,337		
WOODFIELDS	**80.8%**	**97**	**2,434**	**2,433.1**	
SALUDA	81.0%	107	2,541		
MATHEWS	83.3%	84	2,625		
JOHN C CALHOUN	89.9%	36	2,661		
FORD	92.8%	81	2,742		
MIDWAY SCHL	16.7%	142	2,884		
WREN	26.5%	100	2,984		
WRIGHT	28.6%	28	3,012		
POWDERSVILLE	31.2%	173	3,185		
CONCORD	**31.5%**	**133**	**3,318**	**3,194.8**	**Refused**
HUNT MEADOWS	39.8%	75	3,393		
MT LEBANON	41.9%	55	3,448		
MERRIWETHER	48.4%	120	3,568	**3,531.2**	**Replacement**
SPEARMAN	51.3%	60	3,628		
LA FRANCE	51.3%	52	3,680		
BELTON	52.2%	160	3,840		
STARR	53.2%	57	3,897		
CENTERVILLE	**55.5%**	**117**	**4,014**	**3,956.5**	
WEST PELZER	55.6%	69	4,083		
HONEA PATH	55.6%	97	4,180		
CEDAR GROVE	55.9%	90	4,270		
PALMETTO	61.9%	90	4,360		
TOWNVILLE	63.5%	36	4,396		
W E PARKER	64.9%	88	4,484		
MCLEES	65.1%	118	4,602		
IVA	67.2%	70	4,672		
CALHOUN ACADEMY	**67.6%**	**120**	**4,792**	**4,718.2**	
WHITEHALL	69.9%	75	4,867		
NEW PROSPECT	72.3%	66	4,933		
JOHNSTON	73.5%	49	4,982		
HOMELAND PARK	74.7%	52	5,034		
PENDLETON	74.7%	52	5,086		
FLAT ROCK .	76.5%	67	5,153		
NEVITT FOREST	80.7%	66	5,219		
DOUGLAS	82.4%	48	5,267		

Originally selected school Pinecrest refused.

C: New interval start = B Minus Random Start #
909.7 – 148.0 = 761.7

D: New interval size = Sampling Interval Minus A
761.7 – 100 = 661.7

E: Random number between 0 and 1 = 0.17

New selection number = C + (D*E)
761.7 + (661.7 * 0.17) = **874.2**

Replacement school: **Cherokee Trail**

Originally selected school Concord refused.

C: New interval start = B Minus Random Start #
3194.1 – 148.0 = 3046.1

D: New interval size = Sampling Interval Minus A
761.7 – 133 = 628.7

E: Random number between 0 and 1 = 0.56

New selection number = C+(D*E)
3046.1 + (628.7 * 0.56) = 3398.2+133=**3531.2**

Replacement school: **Merriwether**

* If your replacement number is equal to or greater than the refusal school interval, you have to add the enrollment number for the refusing school to your replacement selection number (in this case 133 is added for a final replacement number of 3,533). This adjusts for the fact that the refusing school is no longer in the adjusted interval.

Table 4: Example of refusal replacement for systematic non-PPS sample selection

School Name	NSLP%	3rd Enroll.	Sequential School #	Sampling Interval #	Selected School	Status
KEOWEE	37.1%	38	1	1		
WALHALLA	39.8%	92	2	1		
RAVENEL	52.3%	91	3	1		
LAKEVIEW	52.4%	94	4	1		
NINETY SIX	52.7%	130	5	1		
NORTHSIDE	**52.8%**	**95**	**6**	**1**	**6**	
MCCORMICK	53.9%	56	7	1		
FAIR-OAK	55.7%	117	8	1		
HICKORY TAVERN	56.8%	61	9	1		
HOLLYWOOD	57.2%	66	10	1		
CHEROKEE TRAIL	57.8%	53	11	2		
PINECREST	60.0%	100	12	2		
MERRYWOOD	60.1%	90	13	2		
DIAMOND HILL	60.3%	38	14	2		
SPRINGFIELD	60.9%	89	15	2	**15**	**Replacement**
TAMASSEE-SALEM	**61.7%**	**41**	**16**	**2**	**16**	**Refused**
WESTMINSTER	67.2%	62	17	2		
HODGES	67.5%	36	18	2		
LAURENS	67.7%	92	19	2		
GRAY COURT OWINGS	68.7%	58	20	2		
E B MORSE	68.7%	95	21	3		
CLINTON	69.0%	87	22	3		
WESTWOOD	69.8%	120	23	3		
ORCHARD PARK	71.5%	61	24	3		
WARE SHOALS	72.0%	55	25	3		
JOANNA-WOODSON	**72.2%**	**50**	**26**	**3**	**26**	
JAMES M BROWN	73.9%	98	27	3		
OAKLAND	74.8%	78	28	3		
BLUE RIDGE	76.6%	90	29	3		
WATERLOO	77.5%	37	30	3		
EASTSIDE	78.9%	67	31	4		
WOODFIELDS	80.8%	97	32	4		
SALUDA	81.0%	107	33	4		
MATHEWS	83.3%	84	34	4		
JOHN C CALHOUN	89.9%	36	35	4		
FORD	**92.8%**	**81**	**36**	**4**	**36**	
MIDWAY	16.7%	142	37	4		
WREN	26.5%	100	38	4		
WRIGHT	28.6%	28	39	4		
POWDERSVILLE	31.2%	173	40	4		
CONCORD	31.5%	133	41	5		
HUNT MEADOWS	39.8%	75	42	5		
MT LEBANON	41.9%	55	43	5		
MERRIWETHER	48.4%	120	44	5		
SPEARMAN	51.3%	60	45	5		
LA FRANCE	**51.3%**	**52**	**46**	**5**	**46**	**Refused**
BELTON	52.2%	160	47	5		
STARR	53.2%	57	48	5		
CENTERVILLE	55.5%	117	49	5		
WEST PELZER	55.6%	69	50	5	**50**	**Replacement**
HONEA PATH	55.6%	97	51	6		
CEDAR GROVE	55.9%	90	52	6		
PALMETTO	61.9%	90	53	6		
TOWNVILLE	63.5%	36	54	6		
W E PARKER	64.9%	88	55	6		
MCLEES	**65.1%**	**118**	**56**	**6**	**56**	
IVA	67.2%	70	57	6		
CALHOUN ACADEMY	67.6%	120	58	6		
WHITEHALL	69.9%	75	59	6		
NEW PROSPECT	72.3%	66	60	6		
JOHNSTON	73.5%	49	61	7		
HOMELAND PARK	74.7%	52	62	7		
PENDLETON	74.7%	52	63	7		
FLAT ROCK	76.5%	67	64	7		
NEVITT FOREST	80.7%	66	65	7		
DOUGLAS	**82.4%**	**48**	**66**	**7**	**66**	
VARENNES ACADEMY	90.8%	65	67	7		
SPARTANBURG	43.2%	45	68	7		
LOCKHART	58.8%	23	69	7		
BUFFALO	70.9%	103	70	7		

Selected school Tamassee-Salem refused.

Refusing school is in interval 11-20. Select random number in interval 11-20 = **15**

Replacement is school 15, **Springfield**.

Selected school La France refused.

Refusing school is in interval 41-50. Select random number in interval 11-20 = **50**

Replacement is school 50, **West Pelzer**.

Glossary

Cluster sampling: Sampling groups rather than individuals. This is **practically always** necessary to gain efficiencies in survey logistics, i.e. larger samples can be assessed for lower cost. For example, in surveys of schoolchildren, *multistage cluster sampling* is typically employed, e.g. sampling school districts, then schools within the selected districts, then possibly classes within selected schools. Cluster sampling effectively reduces your sample size to the extent that individuals in clusters are more similar than individuals between clusters, that is, confidence intervals are usually wider **for given sample size** in cluster sampling than if simple random sampling were employed (see Design effect). The specific issue is intracluster or intraclass correlation, which is accounted for in statistical programs designed for analysis of survey data employing complex sample designs (e.g. SUDAAN, and survey analysis procedures in SAS and Stata). The degree of intracluster correlation varies by the outcome analyzed, differing for different outcomes measured in the same survey. An obvious example of cluster sampling in a 3rd grade survey is the selection of schools, with screening of children from selected schools.

Design effect: The ratio of the variance of the estimator based on the employed complex sample design to the variance of the estimator based on a simple random sample of the same size.

Nonresponse: Nonresponse can be associated with nonresponse bias (refusing children differing from participating children in the survey variables of interest), which is difficult to assess and can vary depending on the population and the variable. To the extent that response rates differ from school to school, the ultimate probability of being in the survey sample will differ, even when schools are sampled with equal probability.

Oversampling: Sampling groups at different rates within a population. This is done when you are interested in getting good estimates of a particular subpopulation (e.g. low SES) or when you are particularly interested in estimates and comparisons for subgroups of the population, e.g. different regions of a state, rather than just maximizing precision of overall population estimates. With set resources, oversampling will likely have a small effect on loss of precision (i.e. wider confidence interval) of overall population estimates and estimates of the groups that end up under-sampled. NOTE: this can occur when pursuing county level data in a state survey with limited resources, where the resulting sampling will likely oversample low population rural counties and under-sample high population urban counties.

Probability sampling: Sampling using some form of random selection where every element in the survey population has a known or calculable non-zero chance of selection. Probability sampling is necessary for valid estimates of statistical precision.

Purposive ("expert choice") Sampling: Using knowledge of a population to pick or select the elements for the sample that the sampler thinks represents the target population. This may be preferable to probability sampling where a small number of elements are to be selected or where there is an increased potential for sampling error with random methods. However, purposive sampling can introduce bias and does not allow for statistical precision estimation. *Quota Sampling* is a variation of this, where specified numbers of subcategories of the population are purposively selected in proportion to population numbers in an attempt to achieve better representation of the population.

Stratification: Dividing the population into groups and sampling separately from the groups. This improves your sample to the extent that individuals in the groups are more similar than individuals between groups. Clusters are sampled from all defined strata, ensuring that all strata are represented in the survey sample. Stratification will usually improve precision of population estimates from survey data. Improvements in estimate precision from stratification, however, are usually outweighed by losses in precision due to intracluster correlation associated with cluster sampling.

Weights: Generally the inverse of the probability of selection and survey participation (consented and information collected), which in effect is the number of children in the target population that each person in the sample represents. This direct method is weighting based on sample design and includes adjustment for non-response. Post-weighting can also be employed, adjusting the survey sample to known population parameters.