Sample Selection and Weighting Features of the National Safe Schools/Healthy Students Initiative Evaluation Study

Shelton Jones, Lei Li, Joseph McMichael, Harper Gordek, and Allison Burns
RTI International

Keywords: Survey sampling, sample weighting, multi-wave study

Abstract

Improving school safety and promoting healthy development among school-aged children requires cooperative efforts with schools, communities, families, law enforcement, and other partnering organizations. The Safe Schools/Healthy Students Initiative funded 97 sites composed of one or more Local Educational Agencies nationwide to support local partnership development, safe school police, and program implementation. This evaluation study conducts surveys of students, classroom teachers, principals, school administrators, district coordinators, and other administrators. A sampling frame of virtual classes was constructed based on the Common Core of Data database used in selected study sites and a stratified random sampling method was employed to obtain probability samples. A baseline and two follow-up rounds of surveys were conducted to accommodate longitudinal between year comparisons as well as cross-sectional estimation. This paper discusses characteristics of the sample selection and weighting procedures implemented in the evaluation study.

1. Introduction

The Safe Schools / Healthy Students (SS/HS) Initiative represents the joint efforts of the Federal Departments of Justice, Health and Human Services, and Education to improve school safety, mental health and well-being. Nationwide, 97 Local Education Areas or study sites have been funded to establish comprehensive, integrated strategies to promote the healthy development of students and families in a safe school and community environment. Each site may consist of a partial school district, a single district, or a combination of multiple school districts.

The national evaluation study attempts to assess the effectiveness and overall impact of the Initiative. Major interest includes the effects of the Initiative on the reduction of violent and delinquent behaviors, development of policy and programs, and the formation and functioning of school and community partnerships. The target population is defined as the eligible units receiving funds and participating in the SS/HS research initiative. Eligible units are sites, public school district personnel, public school personnel, classroom teachers, and students. Hence, this study is a multi-survey design and provides adequate statistical power to detect specified levels of changes in survey response variables between baseline measurements and those from a follow-up year.

Probabilistic sample surveys were conducted to provide estimation for each individual selected study site and for the entire SS/HS Initiative Program, that is, for the aggregation of all study sites. Consequently, individual site level statistical analysis weights were developed as the product of the inverse of the stage-wise selection probabilities for different survey types times the sampling unit level nonresponse adjustment(s). Program level analysis weights are products of the site analysis weights, the site level sampling weight, and a post-stratification ratio adjustment as appropriate. The post-stratification included using a Generalized Exponential Model (GEM) approach for sampling weight calibration described by Folsom and Singh (2000).

Although this evaluation study is designed to accommodate multiple surveys, most of the survey samples are based on a single stage of sampling. A stratified random sample of individual classes was selected from a sampling frame of virtual classes. Once the class sample was selected, we determined which schools, school personnel, classes, and students were eligible for participation followed by the districts associated with the eligible schools.

Below, Section 2 provides features of the virtual sampling frame including the sample selection methods, and Section 3 describes sample weighting procedures for the various types of surveys. Also contained in these sections is survey type specific information relative to the procedures for producing weighted individual site estimates and weighted SS/HS Program level estimates.

2. Frame Construction and Sample Selection

A sampling frame of virtual or pseudo classes was constructed from the 1999/2000 Common Core of Data (U.S. Dept. of Education) for school enrollment and from information secured by site team members. These classes were formed based on an estimated 25 students per class within each eligible grade (1, 3, 5, 7, 9, 11). Due to sample size constraints, no sampling was conducted in small sites that is, those sites consisting of fewer than 600 estimated eligible students. This was because limited population counts of teachers and students at small sites would result in substantial reductions in precision of site level estimates if sampled. Therefore, all frame units within the small sites were included in the study with certainty.

For the remaining sites or the large sites, probability sampling was employed. Specifically, a stratified random sample of 17 classes was selected within each eligible grade participating in the large sentinel sites of the student surveys.

Student questionnaires were fielded in Student Surveys (SS) for students of grades 7, 9, and 11, and in Teacher Rating Scales (TRS) for grades 1, 3, and 5. Teacher questionnaires were fielded in Classroom Teacher Surveys (CRT) for participating teachers of elementary grades 1, 3, and 5 followed by the school personnel surveys. In addition, data were collected from district superintendents, project directors, and key partners in mental health, law enforcement and education in the participating sites.

For the classroom teacher survey, a sample of 10 classroom teachers were selected for participation from each elementary grade (1, 3, 5), and 20 classroom teachers were selected in each of the middle and high school grades (7, 9, 11) for a total of 7,244 initial virtual teacher selections. A school associated with a classroom teacher that was selected for the teacher survey was included with certainty in the other school personnel surveys: Principal Survey, Violence Prevention Survey, Mental Health Survey, and Substance Abuse Specialist Survey.

Our sample design reduces the effect of unequally weighting the school personnel sample and benefits from logistical considerations employed to develop the sampling frame and collecting the data. This suggested that a single-stage sample design be used for the classroom teacher survey as well. The clustering of teachers within schools should add marginally to the design effects and thus the sample size needed for the teacher survey, but it greatly enhanced the efficiency of sample frame development and data collection.

The SS/HS Initiative awarded grants to three cohorts of study sites. The first cohort is comprised of 54 initial study sites funded in Fall 1999. Cohorts 2 and 3 are comprised of 23 additional sites funded in Spring 2000 and 20 additional sites funded in Fall 2001, respectively.

This evaluation study uses a multi-wave multi-survey design. A baseline survey and two follow-up surveys were conducted in 2001, 2002, and 2003 for the first two cohorts, and in 2002, 2003, and 2004 for the third cohort. The list of schools was updated by removing closed schools or adding newly opened schools for the follow-up surveys. The student surveys and teacher rating scale surveys used a stratified three-stage sampling design. Twenty sentinel sites were randomly selected from strata formed by region and urbanicity. Classes were then selected in each sentinel site, all students in selected classes were included for student surveys and five students per class were randomly selected for teacher rating scale surveys. The classroom teacher survey sampled classes stratified by site then teachers in the selected classes were included. The school-based personnel surveys did not select a school sample. Instead, all schools whose classes were included in the classroom teacher surveys were included with certainty.

This sample design is similar to the single-stage design used in the national School-based Substance Use Prevention Programs Study (SSUPPS), (Jones et al., 2002). The SSUPPS study involved a single-stage selection of schools. Once schools were selected, research interest also began to focus on school district-level estimation, although districts were never selected. We then derived a multiplicity adjustment estimator (Sirken, 1972) for district weighting as a function of the inverse of the school selection probabilities within each district inclusion. Similarly for this study, classes are our virtual sampling units leading to the selection of students and the inclusion of schools and districts.

3. Weighting the Sample

This study is intended to evaluate the impact of the SS/HS Initiative within each study site as well as the overall program-wide impact covering all 97 sites that have received the SS/HS Initiative funding. As sampling sites was done in student surveys and teacher rating scale surveys, we created two sets of analysis weights, site-level weights and program-level weights, to meet the needs for site-specific analysis and program-wide analysis. These weights were also weighted up for survey nonresponse or post stratified to know margins using GEM or weighting class adjustment method discussed in Kalton and Maligalig (1991), Jones and Chromy (1982), and Chapman (1976). Table 1 shows an example of estimated totals that student-level weights were post stratified to in order to reduce nonresponse bias.

3.1 Student Survey Weights

Let N_h be the number of sites participating in the SS/HS Initiative and n_h be the number of sites selected within each stratum h. The site selection probability is $\pi_{hi} = n_h/N_h$. The student selection probability $ss\pi_{hijklm}$ within class l, grade k, school j, site i and stratum h is a constant by grade within each site, denoted by π_{hik}. Let N_{hik} be the estimated number of classes on the frame within stratum h, site i, and grade k, or the total enrollment divided by 25, and n_{hik} be the number of classes selected, then the student selection probability is n_{hik}/N_{hik}. The site-level weight is simply the inverse of π_{hik}, $ssw_{hijklm} = 1/\pi_{hik}$. And the program-level weight is the inverse of the product of site selection probability and student selection probability. The weights were then adjusted for student nonresponse within responding classes. The classroom size data were collected by the field interviewers using a header form completed by the interviews and then used to determine the adjustment factors for student nonresponse. The adjustment factor for class nonresponse was the ratio of the sum of weights for the eligible (respondent and nonrespondent) classes selected within grade over the sum of weights for the responding eligible classes. The nonresponse adjusted site-level weights and program-level weights were then post-stratified using GEM. The final weights were brought up to the total student counts on the sampling frame by grade, region, and urbanicity as shown in Table 1.

3.2 Teacher Rating Scale Survey Weights

In teacher rating scale surveys, each selected teacher was asked to identify from their student rosters up to five randomly pre-selected students per class and to complete a rating scale questionnaire on each student selected. Similar to the student survey weights, two sets of sample weights for student samples were calculated at the site level and program level.

Let N_{hik} be the total number of classes for the Teacher Rating Scale survey within grade k, site i, stratum h, and n_{hik} be the number of classes selected, the probability of selecting a class or teacher is n_{hik}/N_{hik}. Further, let N_{hikl} be the total number of students in the selected class l and n_{hikl} be the number of students selected, generally, $n_{hikl} = 5$. Then the student selection probability within a site is the product of the class selection probability and the student selection probability,

$$tr\pi_{hijklm} = n_{hik}/N_{hik} * n_{hikl}/N_{hikl}.$$

Hence, the site-level teacher rating scale weight for a student m is $trw_{hijklm} = 1/tr\pi_{hijklm}$. The program-level weight is the inverse of the product of the site selection probability and the student selection probability within a site. Since teachers rated each selected student in a class, a nonresponse adjustment factor was also essential. The adjustment factor was the ratio of the sum of weights for the eligible students within grade k, site i and stratum h over the sum of weights for the students with completed teacher rating data.

3.3 Classroom Teacher Survey Weights and School-Based Survey Weights

The class selection probability for class l in stratum h, site i and grade group k (grades 1, 3, 5 and grades 7, 9, 11) is $\pi_{hikl} = n_{hik}/N_{hik}$, where n_{hik} and N_{hik} are the sampled number of classes and the total number of classes in grade group k. The classroom teacher survey weight is the inverse of the selection probability. These weights were adjusted for nonresponding teachers as well.

Sample weights for schools were derived from the perspective of estimating school-level quantities using the classroom teacher survey data since the inclusion of the school sample depended on the class sample. Let n_{hij} and N_{hij} be the number of classrooms drawn in school j in site i of stratum h in the classroom teacher survey and the total number of classrooms in school j. Clearly, $n_{hi} = \sum_j^{m_{hi}} n_{hij}$ and $N_{hi} = \sum_j^{M_{hi}} N_{hij}$ are the total number of classes sampled in site i and the total population of classes, where m_{hi} and M_{hi} are the number of schools included in the school-based surveys in site i and the total population of schools. We can denote a quantity of interest in the school-based surveys from school j in site i by y_{hij}. We can convert it to be a classroom-level quantity, $x_{hijl} = y_{hij}/N_{hij}$, where l = 1, 2, …, N_{hij}.

Then the total of y_{hij} in site i is $Y_{hi} = \sum_j^{M_i} \sum_\ell^{N_{ij}} x_{hijl}$.

An unbiased estimator of Y_{hi} is as follows.

$$\hat{Y}_{hi} = \frac{N_{hi}}{n_{hi}} \sum_j^{m_{hi}} \sum_l^{n_{hij}} x_{hijl} = \frac{N_{hi}}{n_{hi}} \sum_j^{m_{hi}} \sum_l^{n_{hij}} \frac{y_{hij}}{N_{hij}}$$

$$= \sum_j^{m_{hi}} \left(\frac{n_{hij}}{n_{hi}}\right)\left(\frac{N_{hi}}{N_{hij}}\right) y_{hij}$$

Therefore, the weight for school j in site i can be taken to be $\left(\frac{n_{hij}}{n_{hi}}\right)\left(\frac{N_{hi}}{N_{hij}}\right)$. Note that $\frac{N_{hij}}{N_{hi}}$ would be the probability-proportional-to-size (PPS) selection probability of school j had we drew schools with probability proportional to number of classrooms. Sample weights were also adjusted for nonresponding school personnel within a site. An estimate of N_{hi} was obtained by summing up the classroom teacher survey weights.

3.4 Longitudinal Weights

Although the SS/HS Initiative evaluation study was not designed to be a follow-up study, measuring changes in some performance indicators between the baseline surveys and follow-up waves of surveys is of major concern. The sampling frame used for sample selection varied over time due to the withdrawal of some study sites, school closure, and the opening of new schools. This could affect the comparability of estimates between different waves. We calculated a set of sample weights by restricting the study population to school eligibles for all the waves of the surveys, and considered the sampled schools as if they were followed up over all the waves of surveys.

The longitudinal school-based survey weights were based on the weights for the baseline survey to reflect for changes in school eligibles for the surveys. This was to account for the sampling frame reduction due to school closure and site dropouts. The newly opened schools did not affect the baseline weights and were excluded from the longitudinal samples. In addition, sampled schools were considered nonrespondents if they did not respond in any wave of the survey.

4. Conclusions

Our sampling design was motivated by the charge to provide for estimation at both the program level and for each of the 97 study sites. In addition, our design had to provide estimation for seven surveys of interests: Student Survey, Teacher Rating Scales, Classroom Teacher Survey, and the four school-based surveys, as well as designed to provide for HLM modeling as appropriate. We therefore chose this single-stage design involving the sampling of a virtual frame of classes of students.

5. References

Chapman, D.W. (1976). "A Survey of Nonresponse Imputation Procedures." Proceedings of the *American Statistical Association, Social Statistics Section*, pp. 245-251.

Folsom, R.E, Jr., and A.C. Singh (2000). "The Generalized Exponential Weight Calibration Model for Extreme Values, Nonresponse, and Poststratification." Proceedings of the *American Statistical Association, Section on Survey Research Methods*. This paper was supported by the Substance Abuse and Mental Health Services Administration, Department of Health and Human Services.

Jones, S.M., and J.R. Chromy (1982). "Improved Variance Estimators Using Weighting Class Adjustments for Sample Survey Nonresponse." Proceedings of the *American Statistical Association, Section on Survey Research Methods*, pp. 105-110.

Jones, S.M., B.C. Sutton, and K.E. Boyle (2002). "Survey Methodology for Studying Substance Use Prevention Programs in Schools." In Chaubey, Y.P. (Ed.), *Recent Advances in Statistical Methods: Proceedings of Statistics 2001 Canada*, 4th Conference in Applied Statistics, World Scientific, London.

Kalton, G., and D.S. Maligalig (1991). "A Comparison of Methods of Weighting Adjustment for Nonresponse." *1991 Annual Research Conference Proceedings*, pp. 409-428. U.S. Department of Commerce, Economics and Statistics Administration, Bureau of the Census.

Sirken, M.G. (1972). "Stratified Sample Surveys with Multiplicity." *Journal of the American Statistical Association*, Vol. 67, No. 337, pp. 224-227.

U.S. Dept. of Education, Public Elementary/ Secondary Universe Survey (National Center for Education Statistics).

Table 1. 2001/2002 Estimated Students Used in GEM Post-Stratification for Student Surveys

		Total	7	9	11
Total	**Total**	361932	130739	128151	103043
	Large Urban	143825	49651	54311	39863
	Mid-sized Urban	134182	51566	43298	39318
	Rural	28715	8161	11450	9105
	Suburban	54817	21276	18875	14667
	Tribal	393	86	217	90
Northeast	**Total**	94121	35592	32750	25779
	Large Urban	78703	29081	27776	21846
	Mid-sized Urban	10742	4566	3449	2728
	Rural	2979	1315	921	744
	Suburban	1697	630	605	461
	Tribal	0	0	0	0
Midwest	**Total**	40625	15204	12695	12726
	Large Urban	0	0	0	0
	Mid-sized Urban	24080	8241	7596	8242
	Rural	3106	1035	1042	1029
	Suburban	13323	5928	3983	3412
	Tribal	116	0	74	42
Southern	**Total**	84248	28201	29917	26130
	Large Urban	0	0	0	0
	Mid-sized Urban	78559	26792	27698	24069
	Rural	4827	1133	1834	1860
	Suburban	862	275	386	201
	Tribal	0	0	0	0
Western	**Total**	142939	51742	52788	38409
	Large Urban	65122	20569	26536	18016
	Mid-sized Urban	20801	11967	4555	4280
	Rural	17803	4678	7654	5471
	Suburban	38935	14441	13900	10593
	Tribal	278	86	144	48