DOCUMENT RESUME

ED 216 041 TM 820 204

AUTHOR Benrud, C. H.; Williams, R. L.
TITLE Development of Adjusted School Weights for the National Longitudinal Survey of the High School Class of 1972. Contractor Report.
INSTITUTION Research Triangle Inst., Research Triangle Park, N.C. Center for Sampling Research and Design.
SPONS AGENCY National Center for Education Statistics (ED), Washington, DC.
REPORT NO NCES-82-208
PUB DATE May 81
CONTRACT OE-0-73-6666
NOTE 17p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Data Analysis; Data Collection; Graduate Surveys; High Schools; Research Design; *Research Problems; *Sampling; Statistical Analysis
IDENTIFIERS *National Longitudinal Study High School Class 1972; *Unit of Analysis Problems

ABSTRACT

The major emphasis of the National Longitudinal Study of the High School Class of 1972 (NLS) has been directed toward individuals in the 1972 high school graduating class. Consequently, sampling parameters were established to ensure student representation, and the bulk of data collected were individual-level data. School-level data exist on a separately documented database, and weights appropriate for analyzing these data have been computed. The purpose of this report is to define the post-stratification adjustment procedure used to develop appropriate school-level weights, given an initial school selection procedure designed to provide student representation. This report describes the NLS school sample design, computation of unadjusted student weights (used in determining weights for individual high school seniors), and computation of adjusted weights for school-level analyses. (Author/BW)

Contractor Report

Development of Adjusted School Weights for the National Longitudinal Survey of the High School Class of 1972

National Center for Education Statistics

Development of Adjusted School Weights for the National Longitudinal Survey of the High School Class of 1972

Sampling Research and Design

C. H. Benrud
R. L. Williams

Andrew J. Kolstad
Project Officer
National Center for Education Statistics

May 1981

Prepared for the National Center for Education Statistics under contract OE-0-73-6666 with the U.S. Department of Education. Contractors undertaking such projects are encouraged to express freely their professional judgment. This report, therefore, does not necessarily represent positions on policies of the Government, and no official endorsement should be inferred. This report is released as received from the contractor.

NCES 82-208

ACKNOWLEDGMENTS

The authors wish to thank all those people who have contributed to the preparation of this report. Various constructive comments and suggestions have been provided by the NCES, Project Officer, Andrew Kolstad, by William B. Fetters of NCES, and by Doug Wright of DMES. Dr. G. J. Burkheimer of RTI also provided editorial contributions.

A final word of acknowledgment and an expression of gratitude is due to the many persons in the Federal government and at RTI who assisted in planning and implementing the National Longitudinal Study of the High School Class of 1972; to the more than 20,000 young adults who took the time and effort to provide comprehensive, detailed information about their lives; and to the participating high schools that made it possible to initiate the study in 1972.

TABLE OF CONTENTS

LIST OF TABLES

I. INTRODUCTION

A. Background

The National Longitudinal Study of the High School Class of 1972 (NLS) is a large-scale survey which has been conducted by Educational Testing Service (ETS) and the Research Triangle Institute (RTI) for the National Center for Education Statistics (NCES). The NLS was designed and implemented to allow a better understanding of the development of young adults as they pass through the American educational system and of the complex factors associated with individual educational and career outcomes.

Following a rather extensive period of planning, which included the design and field test of survey instrumentation and procedures, a full-scale survey was initiated in spring 1972. This base-year survey (Hilton and Rhett, 1973), conducted by ETS, collected student data onsite in the high schools, including high school record information, scores on a specially developed test battery, and questionnaire responses about high school experiences, background, attitudes and opinions, and plans for the future. Other information about the high school and its educational programs also was collected in the base year.

Subsequently, four major follow-up surveys have been conducted by RTI from October through April or May of the years 1973-74, 1974-75, 1976-77, and 1979-80. They involved the collection, by mail and personal interview, of student questionnaire data including information about activities and accomplishments since the prior survey, attitudes and opinions, and plans for the future. Additionally, RTI has conducted several supplemental survey efforts to collect, retrospectively, key items of information that had not been provided during prior major surveys. Methodological details of the several follow-up surveys are available elsewhere (Bailey, 1976(b); Bailey, 1976(b); Levinsohn and McAdams, 1978; Riccobono, Burkheimer, and Place, 1981).

B. Purpose of this Report

As indicated above, the major emphasis of NLS has been directed toward individuals in the 1972 high school graduating class. Consequently, major sampling parameters were established to ensure student representation, and the bulk of data collected has been individual-level data. Individual-level data exists on a well documented data base with appropriate adjusted individual sampling weights to allow a large number of individual-level analyses (c.f., Riccobono et al., 1981).

Other data at the high school level were collected, however, as indicated above, and interest has existed in analyzing school level data and/or student data aggregated to the school level. The school-level data exist on a separately documented data base (Henderson and Levinsohn, 1981), and weights appropriate for analyzing these data have been computed and included on the file. The purpose of this report is to define the post-stratification adjustment procedure used to develop appropriate school-level weights, given an initial school selection procedure designed to provide student representation, which included oversampling of some schools and sampling of some others with probability proportional to student population size. The three remaining sections of this report describe (1) the NLS school sample design, (2) computation of unadjusted student weights (used in determining weights for individual high school seniors), and (3) computation of adjusted weights (for school-level analyses.

II. THE NLS SCHOOL SAMPLE

A. Original Sample Design

The sample design may be described as a deeply stratified probability sample of schools. The population consisted of all public, private, and church-affiliated high schools in the 50 States and the District of Columbia. The sampling frame was constructed from computerized school files maintained by the U.S. Office of Education and by the National Catholic Education Association (Westat, Inc., 1972).

The school sampling frame was stratified into 600 final strata based on the following variables:

-Type of control (public or nonpublic),

-Geographic region (Northeast, North Central, South, and West),

-Grade 12 enrollment (less than 300; 300 to 599; 600 or more),

-Proximity to institutions of higher learning,

-Percentage minority group enrollment,

-Income level of the community, and

-Degree of urbanization.

In order to increase the numbers of disadvantaged students in the sample, schools located in low income areas and schools with high proportions of minority group enrollments were sampled at approximately twice the sampling rate used for the remaining schools. Schools in the smallest grade 12 enrollment strata (fewer than 300 seniors) were selected with probabilities proportional to their estimated numbers of senior students and without replacement. Schools in the remaining strata were selected with equal probabilities, without replacement. Within each final stratum, four schools were selected initially and then two of the four were randomly selected and designated as the primary selections. The other two schools were retained as backup or substitute selections and were used in the sample only if one or both of the primary schools did not cooperate (e.g., refused, ineligible).

In order to provide for representation of eligible schools that had not been included in the lists used in making up the sampling frame, a subsample of 200 school districts was selected from among the districts represented in the primary sample of 1,200 schools. These districts were surveyed to obtain names and addresses of all schools with grade 12 enrollment that had not been included on the original sampling frame, and an "augmentation sample" of 23 schools was chosen from the list of 45 schools so compiled.

B. The Final Sample

Of the 1,200 schools that were originally selected, 21 were found to have no 12th grade students and 231 refused, leaving 948 primary schools to participate in the Base Year survey. Of the replacements that were obtained from the backup list, 122 schools participated, yielding a total of 1,070 base-year participating schools. In 26 cases, school-level data were collected from backup schools on the basis of a preliminary refusal of a primary sample school that later agreed to participate. These 26 cases, of which 18 also provided student-level data, thus represented "extra" schools (i.e., more than the required two schools per stratum). None of the augmentation sample schools were included in the Base Year survey (c.f., Hilton and Rhett, 1973).

Because of the large school nonreponse in the Base Year, before beginning the First Follow-up survey a resurvey effort was initiated to elicit responses from the 231 nonparticipating schools and to obtain replacements for all of the 21 schools that had no seniors (and therefore could provide no student-level data). As a result of this effort, 205 of the original refusal schools agreed to participate, together with 35 backup schools. The eight base-year "extra" schools from which no student data were obtained were dropped from the sample, and 16 of the 23 sampled augmentation schools participated, yielding a grand total of 1,318 participating schools in the final school sample. This, of course, represents more than the 1,223 schools that were anticipated in the original design and, therefore, results in strata with more than two schools. Due to the expense involved in collecting student data, however, a post hoc sample redefinition was implemented rather than discarding data already collected (c.f., Moore and Shah, 1975; Moore, 1975; Shah and Marnell, 1976). The number of participating schools in various categories is summarized in Table 1.

TABLE 1. NUMBER OF PARTICIPATING SCHOOLS BY CATEGORY

Category	Base Year Participants	Resurvey and Augmentation Additions	Total Participating
Primary	948	205	1,153
Backup			
Extra[a]	26*	-	18*
Other	96	35	131
Augmentation	-	16	16
Total	1,070	265	1,318

[a] Backup schools from strata in which both primary schools ultimately participated during the base year.

* For 8 of the extra schools no student data were ever collected; consequently, these schools were dropped from the sample.

III. UNADJUSTED SCHOOL WEIGHTS

For each of the 1,302 schools of the original strata (the nonaugmentation schools), the sample inclusion probability, P_{hi}, for selecting school-i in stratum-h was calculated as

$$P_{hi} = n_h A_{hi} / A_h \ ,$$

where n_h = number of participating schools in the final NLS sample for stratum-h,

A_{hi} = size measure for school-i of stratum-h,

and A_h = sum of size measures for all schools in stratum-h.

For nonaugmentation school strata defined by schools with less than 300 seniors, within which PPS sampling had been implemented, the size measure was the larger of the estimated number of seniors in the school and 18. For the remaining nonaugmentation school strata, within which simple random sampling had been implemented, the size measure was one. The use of the number of participating schools in each of the original final strata (n_h) accounts for the strata where the number of participating schools was not equal to two (c.f., Moore, 1975; Shah and Darnell, 1976).

In the sample of 200 school districts canvassed during the Base Year to identify public schools not included on the sampling frame, a total of 45 such schools were identified and 23 were randomly selected as an augmentation sample (Hilton and Rhett, 1973). School selection probabilities, P_i, were calculated for each of the 45 schools, as specified by the NLS sampling contractor (Westat, 1972). Based on a joint NCES and RTI decision, the 16 participating augmentation schools were grouped by pairs into eight additional strata (h = 601, 602, ..., 608), and sample inclusion probabilities, adjusted for school nonparticipation, for these 16 schools were calculated as

$$P_{hi} = \left(\frac{16}{23}\right)\left(\frac{1}{2}\right) P_i \ .$$

For all of the sample schools, unadjusted weights were computed as reciprocals of their probabilities of selection. Explicitly,

$$W_{hi} = [P_{hi}]^{-1} \ .$$

IV. ADJUSTED SCHOOL WEIGHTS

Although the unadjusted weights were appropriate as a base for subsequent calculation of individual-level weights, a post-stratification technique was selected for adjusting the nonaugmentation school weights, so that school-level ratio estimates based on the actual total number of schools in various categories could be routinely produced. This procedure reduces bias in school-level estimates resulting from the irregularities that had entered into the final sample. Post-stratification adjustment forces the weighted distribution of the participating schools to be the same across the post-strata as the population distribution. No adjustments were made to the augmentation school weights beyond that implicit in their inclusion probability formula, since no further information was available for these schools.

Accordingly, weighting classes were established based on six of the primary stratification variables; namely, grade 12 enrollment, geographic region, percentage minority enrollment, income level of the community, type of control and proximity to institutions of higher learning. A separate class was established for the augmentation schools. Within a nonaugmentation school weighting class, the unadjusted weights were summed to obtain an estimate of the number of schools in the class. An adjustment factor was then computed for each class by dividing the actual number of schools listed on the sampling frame in that class by the estimate for that class. The post-stratification adjusted weights were then obtained as the product of the unadjusted weights and the corresponding adjustment factor.

It should be noted that the post-stratification adjustment for schools in the strata with 300 or more seniors (those drawn with equal probabilities within strata) is unity and, thus, has no effect on the unadjusted weights for these schools. This results from the facts that the schools were drawn with equal probabilities within strata and from the way the school inclusion probabilities were defined to account for the use of backup schools. For this reason, all of the large schools were grouped into one weighting class, class 15. The remaining nonaugmentation schools were grouped into the first 14 weighting classes shown in Table 2. The post-stratification adjustment factors, together with the number of sample and frame schools and the sums of unadjusted and post-stratification adjusted weights, are shown in Table 3 for both nonaugmentation schools and augmentation schools (weighting class 16).

TABLE 2. WEIGHTING CLASSES USED IN MAKING POST-STRATIFICATION ADJUSTMENT TO NONAUGMENTATION SCHOOL WEIGHTS

Weighting Class	Grade 12 Enrollment Size Class[1]	Geographic Region[2]	SES Type[3]	Type of Control[4]	Proximity to Inst. of Higher Learning[5]	No. of Sample Schools
1	1	1	1			24
2	1	1	2	1		61
3	1	1	2	2		31
4	1	2	1			63
5	1	2	2	1	1	20
6	1	2	2	1	2	34
7	1	2	2	1	3	44
8	1	2	2	2		26
9	1	3	1		1	30
10	1	3	1		2	69
11	1	3	1		3	119
12	1	3	2			52
13	1	4	1			26
14	1	4	2			42
15	2,3					661

[1]/1 = under 300; 2 = 300-599; 3 = 600 and over.

[2]/1 = Northeast; 2 = North Central; 3 = South; 4 = West (U.S. Census definitions); blank = not used in definition.

[3]/1 = low socioeconomic status; 2 = high socioeconomic status (Based on community income level and percent minority enrollment (see Westat, Inc., 1972); blank = not used in definition.

[4]/1 = public; 2 = nonpublic; blank = not used in definition.

[5]/1 = location in counties (New England only) or SMSAs containing 100 largest central-city school districts; 2 = location in other counties or SMSAs containing a public 2 or 4-year college or university; 3 = location elsewhere; blank = not used in definition.

TABLE 3. WEIGHTING CLASS ADJUSTMENT FACTORS WITH NUMBERS OF FRAME AND SAMPLE SCHOOLS AND SUMS OF SCHOOL WEIGHTS

Weighting Class	No. of Schools in Sample	No. of Schools in Frame	Sum of Unadjusted Weights	Post-Stratification Adjustment Factor	Sum of Adjusted Weights
1	24	284	272.00	1.04336	284.00
2	61	1529	1457.49	1.04907	1529.01
3	31	896	1049.30	0.85390	896.00
4	63	1459	1440.37	1.01294	1459.01
5	20	476	475.24	1.00161	476.00
6	34	1057	1117.05	0.94625	1057.00
7	44	1783	1723.42	1.03457	1783.00
8	26	675	775.51	0.87040	675.00
9	30	364	318.16	1.14409	364.00
10	69	1099	1177.04	0.93370	1099.01
11	119	2422	2358.23	1.02705	2422.01
12	52	1547	1511.80	1.02328	1547.00
13	26	476	346.61	1.37329	476.00
14	42	1670	2076.76	0.80414	1670.00
15	661	3149	3149.00	1.00000	3149.00
16	16		1474.84	1.00000	1474.84

A final adjustment to the weights of the nonaugmentation schools was necessary to account for the fact that the frame included some schools that did not have a 12th grade enrollment. This was accomplished by multiplying each weight by 1,179/1,200, the proportion of the original primary sample with grade 12 enrollment. Table 4 presents the sums of the weights as finally adjusted by the eligibility factor (.9825), together with the minimum and maximum final weights by weighting class.

TABLE 4. WEIGHTING CLASS SUMS OF WEIGHTS AS ADJUSTED FOR ELIGIBILITY, WITH MINIMUM AND MAXIMUM WEIGHTS

Weighting Class	Sum of Final Weights	Minimum Final Weight	Maximum Final Weight
1	279.03	2.08	34.62
2	1502.25	8.64	99.99
3	880.32	2.95	163.13
4	1433.48	3.99	102.32
5	467.67	9.29	118.21
6	1038.50	7.12	125.35
7	1751.80	6.35	205.39
8	663.19	3.61	151.84
9	357.63	4.41	35.30
10	1079.78	3.17	80.65
11	2379.62	4.72	98.53
12	1519.93	6.26	220.90
13	467.67	5.49	88.38
14	1640.78	6.33	151.91
15	3093.92	.98	10.81
16	1474.84	11.50	563.50

REFERENCES

Bailey, J. P. National Longitudinal Study of the High School Class of 1972: First Follow-Up Survey Final Methodological Report. Research Triangle Park, N.C.: Research Triangle Institute, 1976(a).

Bailey, J. P. National Longitudinal Study of the High School Class of 1972: Second Follow-Up Survey Final Methodological Report. Research Triangle Park, N.C.: Research Triangle Institute, 1976(b).

Henderson, L., and Levinsohn, J. R. National Longitudinal Study of the High School Class of 1972: School File Documentation. Research Triangle Park, N.C.: Research Triangle Institute, June 1981.

Hilton, T. L., and Rhett, H. The Base-Year Survey of the National Longitudinal Study of the High School Class of 1972. Princeton, N.J.: Educational Testing Service, 1973.

Levinsohn, J. R., and McAdams, K. C. National Longitudinal Study: Third Follow-Up Survey Final Methodological Report. Research Triangle Park, N.C.: Research Triangle Institute, 1978.

Moore, R. P. Calculation of Nonresponse-Adjusted Student Weights for Respondents to the NLS Base-Year and First Follow-Up Surveys. Research Triangle Park, N.C.: Research Triangle Institute, 1975.

Moore, R. P., and Shah, B. V. Choosing the "Optimum" Design for an NLS Cohort. Research Triangle Park, N.C.: Research Triangle Institute, 1975.

Riccobono, J. A., Burkheimer, G. J., and Place, C. National Longitudinal Study: Fourth Follow-Up Final Methodological Report. Research Triangle Park, N.C.: Research Triangle Institute, 1981.

Riccobono, J. A., Henderson, L. B., Burkheimer, G. J., Place, C., and Levinsohn, J. R. National Longitudinal Study: Base-Year (1972) through Fourth Follow-Up (1979) Data File Users Manual (three volumes). Research Triangle Park, N.C.: Research Triangle Institute, 1981.

Shah, B. V., and Marnell, D. K. Calculation of Nonresponse-Adjusted Student Weights for Respondents to the NLS Second Follow-Up Survey. Research Triangle Park, N.C.: Research Triangle Institute, January 1976.

Westat, Inc. Sample Design for the Selection of a Sample of Schools with Twelfth-Graders for a Longitudinal Study. Rockville, Maryland: Author, 1972.