

NLS:72 Common Variables Public Dataset

Dataset Version: 2025-10-30

Introduction.....	1
Sampling Frame.....	1
Residential History.....	3
Demographics.....	4
Family Socioeconomic Status.....	4
High School Experience.....	6
Educational Attainment.....	7
Post-secondary Education.....	8
Employment.....	8
Health.....	8

Introduction

EdSHARe's National Longitudinal Study of the Class of 1972 (NLS:72) common variables public dataset gathers together several of the most common variables (derivable from the NLS:72 public data) that we use in EdSHARe analyses. To the extent possible, the definitions of these variables are parallel to those in EdSHARe's [High School & Beyond Study of 1982 \(HS&B:80\) common variables public dataset](#).

The operationalizations used for this dataset constitute default choices, but you should always consider whether the given operationalization is most appropriate for your specific analysis. If your analysis requires a different operationalization, go back to the original data and construct your own variable, as appropriate for your analysis. Refer to the [code](#) for the NLS:72 common variables public dataset as a starting point.

Sampling Frame

NLS:72 had a two-stage stratified sampling design in which high schools were selected within strata, then high school seniors were sampled from each selected school. The

stratification design was complex, including 600 strata from an incomplete cross of: type of control (public or nonpublic), geographic region, grade 12 enrollment (less than 300, 300-599, 600 or more), proximity to institutions of higher learning, percent minority group enrollment, community income level, degree of urbanization. The variable STRATUM indicates the original first-stage sampling strata, which may be used for design-based variance estimation. Additionally, the dataset includes variables related to the stratification components, which may be useful for imputation and response propensity modeling.

The 1986 fifth follow-up subsampled the original NLS:72 sample to reduce survey costs and maximize response rates. We refer to the NLS:72/86 subsample as the “NLS:72 panel,” and this is the base sample used for the NLS:72/25 follow-up. The variable RAWWT provides sampling weights that incorporate all levels of selection for the panel.

Variable	Source	Original Coding	New Coding
STRATUM	FINSTNO		
hsstate	STCODE	(Custom)	(FIPS)
hsregion	STCODE		1 Northeast 2 South 3 Midwest 4 West
hstype	PUBPRIV	0 {MISSING} 1 PUBLIC 2 PRIVATE 3 CATHOLIC	1 Public 3 Private 2 Catholic
hspub	PUBPRIV	0 {MISSING} 1 PUBLIC 2 PRIVATE 3 CATHOLIC	1 Public 0 Not public 0 Not public
hsg12	N_SENRS	(counts)	
hsprox	COLLPROX	1 CITY SCL 2 PUB COLL 3 OTHER 98 {MISSING}	

hsmin	MINRTY	1 <5% 2 5-9% 3 10-19% 4 20-39% 5 40-59% 6 60-79% 7 80% + 98 {MISSING}	
hsincome	INCOME	(11 levels) 98 {MISSING}	
hsurb ¹	SCHQ40	1 RURAL 2 SMALL CITY 3 MEDIUM CITY 4 SUB MED CITY 5 LARGE CITY 6 SUB LARGE CITY 7 HUGE CITY 8 SUB HUGE CITY 98 {MISSING} 99 {LEGIT SKIP}	3 Rural 1 Urban 1 Urban 2 Suburban 1 Urban 2 Suburban 1 Urban 2 Suburban
hscomsz	COMMUNIT72 (SCHQ40)	1 SMALL 2 MEDIUM 3 LARGE 4 VERY LARGE 95 {UNCODEABLE VERBATIM} 98 {ILLEGITIMATE SKIP}	1 Small 2 Medium 3 Large 4 Very large

Residential History

Five-digit zip codes are provided from the 1973, 1974, 1976, and 1979 follow-ups.

¹ Note: The variable used for stratification was likely DEGURBAN in the HSCHOOL data file, but hsurb is based on SCHQ40, which has a more convenient coding for substantive analysis.

Demographics

Variable	Source	Original Coding	New Coding
sex	CSEX72, TFUSEX	1 MALE 2 FEMALE	1 Male 0 Female
race	CRACE FI97, FI98	1 AMERICAN INDIAN 2 BLACK 3 MEXICAN-AMERICAN 4 PUERTO RICAN 5 LATIN-AMERICAN 6 ASIAN-AMERICAN 7 WHITE 8 OTHER	4 Other 2 Black 3 Hispanic 3 Hispanic 3 Hispanic 4 Other 1 White 4 Other ²
marital79	FT165	1 NVR MARRIED-PLAN TO 2 NVR MARRIED-NO PLAN 3 DIVORCED,WIDOWED,SEP 4 MARRIED,FIRST SPOUSE 5 REMARRIED	3 Never married 3 Never married 2 Previously married 1 Married 1 Married
marital86	FI77	1 DIVORCED 2 WIDOWED 3 SEPARATED 4 PSEUDO-MARRIED 5 MARRIED 6 NOT MARRIED	2 Previously married 2 Previously married 2 Previously married 1 Married 1 Married 3 Never married

Family Socioeconomic Status

Parental education is taken as the higher of father's and mother's education. Additionally, NLS:72 included the education level of the sample member's oldest sibling.

Occupations in the NLS:72 historical data are coded according to the [1970 Census occupation classification system](#). NLS:72 includes a prestige-like occupational socioeconomic index attributed to Otis Dudley Duncan (see Appendix K in the original data

² Following the HS&B:80 composite, the "Other" category also includes a small number of cases for which race/ethnicity could not be determined from CRACE and FI97/98.

documentation, [ICPSR 08085](#)). In the common variables dataset, parental occupation is coded as the higher of father's and mother's Duncan score.

Family income in 1972 was recoded using the midpoints of the response option ranges. To obtain a value for open ended top category, the frequencies in each category were fit with a Normal distribution using the cut points indicated by the category ranges. Then, the median of the upper category obtained from a truncated Normal distribution using the fitted mean and standard deviation.

NLS:72 additionally included a set of ten household items that indicate socioeconomic status. The estimated IRT score based on these ten items is included in the common variables dataset as a supplementary socioeconomic status index.

Variable	Source	Original Coding	New Coding
paredu	CFAED86 CMOED86	1 LESS THAN HIGH SCHOO 2 HIGH SCHOOL 3 SOME COLLEGE 4 4-YEAR COLLEGE 5 MA, PHD	1 HS or less 1 HS or less 2 Some college 4 Four-year Degree 5 Graduate Degree
sibedu	BQ90C	1 DOESN^T APPLY 2 DID NOT FINISH H.S. 3 FINISHED H.S OR EQIV 4 ADULT ED PROGRAM 5 BUSINESS, TRADE SCHL 6 SOME COLLEGE 7 COLLEGE GRAD 8 SOME GRAD, PROF SCHL 9 GRAD., PROF. DEGREE	0 Not applicable 1 HS or less 1 HS or less 1 HS or less 2 Some college 2 Some college 4 Four-year Degree 5 Graduate Degree 5 Graduate Degree
parsei	SEIBQ97, SEIBQ101 BQ25B, BQ25C	(see Appendix K)	(Duncan socioeconomic index score)
income72	BQ93	1 LESS THAN \$3,000 YR 2 \$3,000-\$5,999 A YR 3 \$6,000-\$7,499 A YR 4 \$7,500-\$8,999 A YR 5 \$9,000-\$10,499 A YR 6 \$10,500-\$11,999 A YR	1500 <\$3k 4500 \$3-6k 6750 \$6-7.5k 8250 \$7.5-9k 9750 \$9-10.5k 11250 \$10.5-12k

		7 \$12,000-\$13,499 A YR 8 \$13,500-\$14,999 A YR 9 \$15,000-\$18,000 A YR 10 OVER \$18,000 A YEAR	12750 \$12-13.5k 14250 \$13.5-15k 16500 \$15-18k 20250 >\$18k
homesei72	BQ94B-BQ94K		(IRT score)
siblings	SQ119-SQ120		(top coded at 5)

High School Experience

NLS:72 does not include indicators of specific math courses taken, but does include the number of semesters of math and science courses taken.

Variable	Source	Original Coding	New Coding
hstransfer	SRFQ3	1 NO 2 YES	0 Did not transfer 1 Transferred
hsprog	BQ2	1 GENERAL 2 ACADEMIC,COLL PREP 3 VO-TECH, AGRICULTURE 4 VO-TECH,BUSINESS,COM 5 VO-TECH,DISTRPIVE ED 6 VO-TECH, HEALTH 7 VO-TECH HOME EC 8 VO-TECH,TRDE,INDSTRL	1 General 2 Academic 3 Vocational 3 Vocational 3 Vocational 3 Vocational 3 Vocational 3 Vocational
hsmath	SRFQ4CC		(count of semesters)
hssci	SRFQ4A		(count of semesters)
gpa12	BQ5	1 MOSTLY A 2 HALF A-B 3 MOSTLY B 4 HALF B-C 5 MOSTLY C 6 HALF C-D 7 MOSTLY D 8 BELOW D	4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.0

Educational Attainment

Measurement of educational attainment in NLS:72 is substantially more straightforward than in HS&B:80, as there are self-reported highest-degree earned items in each follow-up starting with NLS:72/76 (four years after high school senior year). The self-report items do not include a category for associate's degree attainment, but attainment of two-year degrees can be ascertained from post-secondary transcripts.

Variable	Source	Original Coding	New Coding
edattain76	TQ49A, TQ48CA, TQ48DA	1 HIGH SCHOOL 2 VO, TRADE OR BUS LT2 3 VO, TRADE OR BUS GE2 4 LT TWO YR COLL 5 GE TWO YR COLL TD48CA == 3 TQ48DA == 4 6 FINISHED COLL 7 MASTERS 8 PHD OR PROF DEGREE	1 HS or less 2 Some college 2 Some college 2 Some college 2 Some college 3 Two-year Degree 4 Four-year Degree 5 Graduate Degree 5 Graduate Degree
edattain79	FT67 FT76CA FT76DA	0 DOES NOT APPLY 1 SOME, LESS THAN 2 YRS 2 2 YEARS OR MORE FT76CA == 3 FT76DA == 4 3 FINISHED COLLEGE 4 MASTER'S DEGREE 5 PHD, ADV PROF DEGREE	1 HS or less 2 Some college 2 Some college 3 Two-year Degree 4 Four-year Degree 5 Graduate Degree 5 Graduate Degree
edattain86	FI29	1 SOME HIGH SCHOOL 2 FINISHED HS OR GED 3 VOC-BUS SCH - < 2 YR 4 VOC-BUS SCH - > 2 YR 5 COLLEGE - < 2 YRS 6 COLLEGE - > 2 YRS 7 FINISHED COLLEGE 8 MASTERS DEGREE 9 PHD MD OTHER GRAD	1 HS or less 1 HS or less 2 Some college 2 Some college 2 Some college 2 Some college 4 Four-year Degree 5 Graduate Degree 5 Graduate Degree

Post-secondary Education

Due to the simpler measurement of educational attainment, far fewer auxiliary post-secondary education variables are needed for NLS:72, compared with HS&B:80. Nevertheless, indicators of post-secondary attendance in 1973 and 1974, as well as the post-secondary transcripts, can assist with imputation of missing attainment.

Variable	Source	Coding
pse73 pse74	FQ23, FQ26B, FQ32B, FQ40B, FQ42 SQ9, SQ12, SQ34, SQ66*	1 No PSE 2 Some college 3 Working toward Bachelor's degree 4 Earned BA
psetrans	PETSTRN: DEGREE1-DEGREE3	1 HS or less 2 Some college 3 Two-year Degree 4 Four-year Degree 5 Graduate Degree

Employment

Variable	Source	Coding
employ79 employ86	ACT579 FI3A	0 Not employed 1 Employed
occsei79 occsei86	SEIF12 FI7A	(Duncan socioeconomic index score, see Appendix K)

Health

Variable	Source	Original Coding	New Coding
physdisab	BQ83	1 NO	0 No limiting

		2 YES	physical condition 1 Limiting physical condition
earlyhealth	BQ17K	1 NOT AT ALL 2 SOMEWHAT 3 A GREAT DEAL	1 Not at all 2 Somewhat 3 A great deal