

2023

Add Health Waves I & II Documentation



Report prepared by

Gabriel Schwartz

Wave I & II School Desegregation Disparities



CAROLINA POPULATION CENTER | CAROLINA SQUARE - SUITE 210 | 123 WEST FRANKLIN STREET | CHAPEL HILL, NC 27516

Add Health is supported by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations.

Acknowledgments

This dataset was created through work led by Dr. Rita Hamad (Associate Professor & Director, Social Policies for Health Equity Research Program [SPHERE], University of California San Francisco [UCSF]) and Dr. Gabriel Schwartz (Postdoctoral Fellow, Institute for Health Policy Studies, UCSF), along with other SPHERE research team members and Dr. Kiarri N. Kershaw and Cyanna C. McGowan at Northwestern University. Dr Hamad’s work was done in partnership with faculty and research staff from the National Longitudinal Study of Adolescent to Adult Health (Add Health) at the Carolina Population Center (CPC) at the University of North Carolina Chapel Hill (UNC). As part of the construction of this data, Dr. Hamad and her team at UCSF extended the research originally done by Dr. Sean Reardon (Director, Educational Opportunity Project, Stanford University) and his research team at Stanford University. Supplemental data was also provided by ProPublica, a non-profit journalism organization. Funding and support for this project was provided through a grant from the NIH National Heart, Lung, and Blood Institute (R01HL151638), and grants from the Research Evaluation and Allocation Committee and the Huntington Fund at the University of California San Francisco.

Introduction

SCHDSG12.xpt contains data on the levels of school racial segregation experienced by Add Health respondents during their school-age years, related school district characteristics, and measures of tract-level residential segregation present in adulthood (Waves III-V). The data set includes:

Demographics Data: Various school district demographics data at Waves I and II. These include number of schools and total number of students enrolled in each district, racial demographics (consisting of percentage make-up of Black, White, and Hispanic students in the district), and the percentage of students eligible for free or reduced-price lunches. Also included in this group are two variables indicating respondent’s Census Region location at Wave I and II as well as a variable tracking whether students switched districts at some point between Waves I and II.

Court Ordered Desegregation Mandates: Which school districts had ever been placed under court desegregation orders mandating them to take active steps towards racial integration, and when some of these orders were subsequently lifted. This group of data consists of separate variables for Waves I and II. These variables detail the calendar year that a district was released from court ordered desegregation mandates as well as the “first fall” a school district was no longer under a court mandated desegregation order (i.e., the first school year). In addition to containing information relevant to the dismissal year of the district, the “Year Dismissed” variables (SDSD01W1 and SDSD01W2) also contain special coding for describing districts that were either never under a court desegregation order or have yet to be dismissed from their court ordered desegregation mandate.

Calculated Measurements of Racial Segregation: A collection of calculated measures of segregation of the student’s school district (as measured at Waves I-II) and residential area segregation (with measurements at Waves I-V). The district level data consists of calculated measurements of “Exposure” and “Dissimilarity” Indexes at Waves I-II measuring segregation at the student’s attended school district, and Local Getis-Ord G^* (GSTAT) statistics measuring neighborhood-level racial residential segregation of Black Americans from other racial/ethnic groups at the time of

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

each wave of Add Health interviews. More detail about these data can be found below in the *Data Structure and Form* section.

Data Structure and Form

Variables related to court-mandated desegregation orders were constructed by Dr. Rita Hamad and Dr. Gabriel Schwartz at UCSF, extending prior work done by the research team led by Dr. Sean Reardon at Stanford University, with supplemental data provided by ProPublica. School district characteristics were created using data from **(1)** the Education Demographic and Geographic Estimates (EDGE) project of the American Community Survey and **(2)** the Common Core of Data (CCD). These characteristics were merged onto the full in-school sample of Add Health at Wave I and onto Waves I and II for the longitudinal sample. Block-level Census “Federal Information Processing Series” (FIPS) codes [hereafter referred to as “block IDs”] for both (1) Wave I and II residential addresses of Add Health participants and (2) Wave I school address enabled the linkage of Add Health sample members to school district data. An explanation of the Census FIPS IDs can be found here [<https://www.census.gov/programs-surveys/geography/guidance/geo-identifiers.html>]. An explanation of school district LEAID codes can be found here [<https://nces.ed.gov/pubs2017/2017074.pdf>].

1. Court Desegregation Orders

In 1954, the Supreme Court issued their landmark *Brown v. Board of Education* decision, which found school segregation unconstitutional. Subsequent decisions followed (*Green v. County School Board of New Kent*, 1968; *Swann v. Charlotte-Mecklenburg Board of Education*, 1971), opening the door for racial justice organizations and families of affected students to bring school districts to court to enforce students’ right to equal educational opportunities.¹ Many of these cases yielded court-ordered or court-enforced desegregation plans, leading to strong declines in school racial segregation.²

Districts could be released from these orders, since orders were not necessarily meant to be permanent. Under *Green v. County School Board of New Kent* (1968), districts could be released if they met seven standards that indicated to the court that desegregation orders were no longer necessary, and that school segregation in their district had been eliminated. In the 1990s, however, *Green’s* standards were relaxed, first in the Supreme Court’s 1991 decision in *Board of Education v. Dowell* and then in *Freeman v. Pitts* (1992) and *Missouri v. Jenkins* (1995). These decisions allowed many affected parties (e.g., parents, district officials) to petition the court to release districts from desegregation orders and made it easier for districts to clear judicial hurdles.^{3,4}

The first set of ancillary variables describe the schooling experiences of students with respect to school district desegregation orders. We construct these variables based on a national database of such orders; this database was first compiled by Dr. Sean Reardon and his team at Stanford University in 2009. It was then extended, first by ProPublica (through 2014) and then by Dr. Reardon’s team again (through 2021). Both Reardon’s 2009 files and ProPublica’s databases are publicly available for free online, although as of the writing of this guide, the 2021 extension may not yet be publicly released.^{5,6} Interested users should check the Stanford Center for Education Policy Analysis website for more recent updates as they are posted.

Variables available to Add Health users include:

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

1. *SCSD01W1* and *SCSD01W2* describe the status of school districts ever under a court desegregation order. They identify school districts, represented in Waves I and II, that were ever under a court desegregation order; if they were released from such an order, the year of release is specified. These variables have been coded to provide a robust measure of districts' dismissal status, offering researchers flexibility in their handling of these records.
2. Variables *SCSD02W1* and *SCSD02W2* report the first school year or "first fall" that order releases went into effect. Though useful, this information could not be identified for all districts and, therefore, the effective fall variables serve to complement *SCSD01W1* and *SCSD01W2*.

SCSD01W1; SCSD01W2	Year district in Wave I or Wave II dismissed
<i>Value</i>	<i>Meaning</i>
<1964-2021>	Year district was dismissed
88880	Districts were never under a court desegregation order
88881	Not dismissed (through 2021)
88882	Dismissed before 1990; date uncertain
88885	Status and/or dismissal date unknown
88886	Not dismissed by 2020; not under order by 1991

2. Special Block ID missing-ness and school district mobility codes

Add Health respondents were occasionally missing residential block ID information at Wave I or Wave II. For such cases the assumption was made that block IDs remained the same and could be substituted as needed for missing values.

Since other researchers may prefer a different method for handling missing block IDs, relevant missing data codes for the variable **SCSD15** are provided and take the following values:

SCSD15	Whether respondents changed school districts between Wave I and Wave II
<i>Value</i>	<i>Meaning</i>
0	No change
1	Changed districts
81	Missing Block ID for Wave I
82	Missing Block ID for Wave II
84	Missing LEAID for Wave II
85	Missing LEAID for both waves
92	Not interviewed at Wave II

3. Measures of Segregation at the School District Level

Measures of segregation were calculated using data from the Education Demographic and Geographic Estimates (EDGE) project of the American Community Survey and the Common Core of Data (CCD) and describe the levels of segregation children experienced in the post-*Board of Education v. Dowell* era (i.e., from 1991 on). Relevant variables available in this file describe levels of school and residential racial segregation at the school district level. These include both school Black-White segregation between schools (i.e., schools are the units within which students are segregated) and residential Black-White segregation between Census tracts (i.e., tracts are the units within which students are segregated).

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

Black-White school segregation exposures across students' school-age years were summarized for Add Health and stored in several variables:

- The level of school segregation within students' school districts at Waves 1 and 2 (SCSD03W1, SCSD03W2, SCSD04W1, SCSD04W2).
- The average level of school segregation experienced between 1991 and each of those waves (SCSD07W1, SCSD07W2, SCSD08W1, SCSD08W2).
- The average level of school segregation experienced between 1991 and their expected 12th grade school year, based on the grade and year of their Wave I interview (SCSD16, SCSD17).

Students are assumed to have attended the same school districts between 1991 and the Wave I 1994-1995 interview and are assumed to have attended the same school districts between Wave II and the expected school year of their high school graduation. (Some students also changed districts between the two waves, which is accounted for—see Section 2.) School district-level residential segregation is captured by variables recording levels of residential segregation in the most recent Decennial Census (1990) before Add Health Waves I and II (SCSD05W1, SCSD06W1).

In a handful of districts with low numbers of Black or White students, segregation measures were not available for every year. This is driven by small cell sizes. When 0 Black or 0 White students attend a given district in a given year, or when counts are unavailable because of small cell suppression, some measures of segregation are incalculable and assigned the code 95 for missing. Segregation exposures averaged across multiple years (e.g., SCSD17) are assigned special missing codes for individual Add Health respondents if data for their school districts were not available for all years in question.

For each school district segregation measure described above, two measures of segregation are available: the *Dissimilarity Index* and the *Exposure Index*.^{7,8} These are described briefly below. Because some of the variables created are averages across the course of a student's school-age years, an indicator for whether students moved from one school district to another between Wave I and Wave II is provided (see Section 2).

The Dissimilarity Index

The Dissimilarity Index represents how evenly two groups are distributed across units of a larger whole. Here, it measures how evenly Black and White students were distributed across schools (or neighborhoods) within a school district. Ranging from 0 to 1, it can be interpreted as the proportion of students in these groups who had to change schools (or neighborhoods) to achieve an even distribution of Black and White students within a district.

For measuring Black-White school segregation, the Dissimilarity Index is here calculated as:

$$D = \frac{1}{2} \sum_{i=1}^N \left| \frac{b_i}{B} - \frac{w_i}{W} \right|$$

where i indexes schools (out of N total schools in a district), b_i is the number of Black students in school i , B is the total population of Black students in the district, w_i is the number of White students in school i , and W is the total population of White students in the district.

The Exposure Index

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

The Exposure Index, sometimes called the “Interaction Index,” represents the extent to which one group is exposed to another within units of a larger whole. Here, using the weighted Black student population, it measures the probability that another student within a Black student’s school is White, ranging from 0 to 1. It is calculated as:

$$X = \sum_{i=1}^N \left(\frac{b_i}{B} * \frac{w_i}{t_i} \right)$$

where i indexes schools (out of N total schools in a district), b_i is the number of Black students in school i , B is the total population of Black students in the district, w_i is the number of White students in school i , and t_i is the total number of students enrolled in school i .

The variables describe Black-White segregation because this is the type of integration that *Brown v. Board of Education* desegregation orders were particularly focused on reducing. Consequentially they may be less relevant for other racial/ethnic groups. Future authors could incorporate additional measures describing the segregation of other racial/ethnic groups from one another, instead of relying on these Black-White segregation variables.

4. Other School District Characteristics

To describe demographic characteristics of school districts at the time the Supreme Court ruled districts could be released from their court-ordered desegregation plans, several variables using the Common Core of Data (CCD) were constructed. These describe the characteristics of districts students were attending at the time of their Wave I and Wave II interviews. District variables include:

- Number of schools (SCSD09W1, SCSD09W2)
- Total student enrollment (SCSD14W1, SCSD14W2) of each district
- Racial composition of its students (SCSD10W1, SCSD11W1, SCSD12W1, SCSD10W2, SCSD11W2, SCSD12W2)
- Proportion of its students eligible for the federal free and reduced-price lunch program (SCSD13W1, SCSD13W2)
- Census region SCSR01W1 and SCSR01W2

5. Neighborhood-level racial residential segregation

Neighborhood-level racial residential segregation of Black Americans from other racial/ethnic groups at the time of each wave of Add Health interviews (Waves I-V) is available as Local Getis-Ord G^* statistics (SCST01W1, SCST01W2),^{9,10} calculated by members of the research team (Dr. Kiarri N. Kershaw and Cyanna C. McGowan at Northwestern University). This G^* statistic is effectively a Z-score: it compares the proportion of the population in a focal tract and its neighboring tracts that is Black to the proportion of a larger surrounding geographic area that is Black. This larger area is a Core-Based Statistical Area (CBSA) for most tracts in Add Health; counties were used for tracts that were not in CBSAs. Any tract sharing an edge or node with the focal tract was considered a neighboring tract (i.e., we used queen spatial contiguity). To interpret a G^* value more concretely, a G^* statistic of 1.96 would indicate a tract and its neighboring tracts have a higher proportion of residents who are Black than the average within their CBSA or county – roughly two standard deviations above the mean.

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

This ancillary data provides neighborhood-level G* statistics using data from the 1990, 2000, and 2010 Decennial Censuses and the American Community Survey. For Waves I and II, G* statistics are based on the 1990 Census; for Wave III, the 2000 Census; for Wave IV, the 2010 Census; and for Wave V, 2014-2018 5-year estimates from the ACS. These G* values are only available for the longitudinal sample (prefixed by “SCST01W_”), as residential block IDs were not available for the in-school-only sample.

G* values were calculated as follows:

$$G_i^* = \frac{\sum_{j=1}^n w_{i,j} x_j - \bar{X} \sum_{j=1}^n w_{i,j}}{S \sqrt{\left[\frac{n \sum_{j=1}^n w_{i,j}^2 - \left(\sum_{j=1}^n w_{i,j} \right)^2}{n-1} \right]}}$$

where x_j is the proportion of tract j 's population who was Black, $w_{i,j}$ is the spatial weight between tracts i and j , n is equal to the total number of tracts, \bar{X} is the mean proportion Black across all tracts, and:

$$S = \sqrt{\frac{\sum_{j=1}^n x_j^2}{n} - (\bar{X})^2}$$

Data Dictionary

Detailed variables, codes and frequency counts are in the SCHDSG12 Codebook. Excepting the unique study ID variable, all variables in this dataset SCHDSG12 follow a defined set of naming schema logic. Each variable (except for the three described further below) is assigned an eight-character variable name. The component parts are:

- (1) “SCS” prefix
- (2) A 3-character location/grouping code: first character identifying the location level of the data (e.g., R=Region level, D=District level, and T=Tract level), followed by a two-digit variable number.
- (3) A final two-character temporal component indicating the Add Health wave number at which this data corresponds to (W1=Wave I, W2=Wave II, etc., continuing through Wave V).

Similar variables that differ only in their temporal component (i.e., what wave number corresponds) are grouped together by this 3-digit letter + number. For example, two variables, SCSD09W1 and SCSD09W2, describe the “number of schools in the district,” with one measured at Wave I and the other at Wave II. Since these variables contain similar information (number of schools) and only differ in the temporal component when they were assessed (the survey wave), their three-digit middle code “D09,” are identical. This “grouping” logic for similar variables extends across all the variables in this dataset.

Note: There are 3 six-character variables in this dataset that do not contain a final two-digit temporal component since their time frame of reference extends across multiple waves. These three variables are SCSD15 (whether participant moved school districts between Wave I and Wave II), SCSD16 (average Black/White district-level school segregation between 1991 and high school graduation (Exposure Index)), and SCSD17 (average Black/White district-level school segregation between 1991 and high school graduation (Dissimilarity Index)).

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

Each variable, along with a brief description, can be found in the table below:

Variable Name	Variable Description
AID	Ancillary study linking ID
SCSR01W1	Census region of students' school district (Wave I)
SCSD01W1	Year district dismissed from court order (for Wave I school districts)
SCSD02W1	The first fall districts were no longer subject to a desegregation order (Wave I school districts)
SCSD03W1	Black/White district-level school segregation at the Wave I interview (Dissimilarity Index)
SCSD04W1	Black/White district-level school segregation in the district students were attending at the Wave I interview (Exposure Index)
SCSD05W1	District-level residential (between tract) segregation at Wave I interview (as measured by the Black-White Dissimilarity Index)
SCSD06W1	District-level residential (between tract) segregation at the Wave I interview (as measured by the Black-White Exposure Index)
SCST01W1	Census tract residential segregation in the 1990 Decennial Census in respondents' Wave I tract (measured as a G* statistic)
SCSD07W1	Average Black/White district-level school segregation between 1991 and Wave I (Exposure Index)
SCSD08W1	Average Black/White district-level school segregation between 1991 and Wave I (Dissimilarity Index)
SCSD09W1	Number of schools in Wave I district in year of interview
SCSD10W1	Percent White students in Wave I district in year of interview
SCSD11W1	Percent Black students in Wave I district in year of interview
SCSD12W1	Percent Hispanic students in Wave I district in year of interview
SCSD13W1	Percent students eligible for free/reduced price lunch in Wave I district in year of interview
SCSD14W1	Total district enrollment in Wave I district in year of interview
SCSR01W2	Census region of students' school district (Wave II)
SCSD01W2	Year district dismissed from court order (for Wave II school districts)
SCSD02W2	The first fall districts were no longer subject to a desegregation order (for Wave II school districts). Year district dismissed from court order (for Wave II school districts)
SCSD03W2	Black/White district-level school segregation at the Wave II interview (Dissimilarity Index)
SCSD04W2	Black/White district-level school segregation in the district students were attending at the Wave II interview (Exposure Index)
SCST01W2	Census tract residential segregation in the 1990 Decennial Census in respondents' Wave II tract (measured as a G* statistic)
SCSD07W2	Average Black/White district-level school segregation between 1991 and Wave II (Exposure Index)

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

Variable Name	Variable Description
SCSD08W2	Average Black/White district-level school segregation between 1991 and Wave II (Dissimilarity Index)
SCSD09W2	Number of schools in the Wave II district in year of interview
SCSD10W2	Percent White students in the Wave II district in year of interview
SCSD11W2	Percent Black students in the Wave II district in year of interview
SCSD12W2	Percent Hispanic students in the Wave II district in year of interview
SCSD13W2	Percent students eligible for free/reduced price lunch in the Wave II district in year of interview
SCSD14W2	Total district enrollment in the Wave II district in year of interview
SCSD15	Whether participant moved school districts between Wave I and Wave II
SCSD16	Average Black/White district-level school segregation between 1991 and high school graduation (Exposure Index)
SCSD17	Average Black/White district-level school segregation between 1991 and high school graduation (Dissimilarity Index)
SCST01W3	Census tract residential segregation in the 2000 Decennial Census in respondents' Wave III tract (measured as a G* statistic)
SCST01W4	Census tract residential segregation in the 2010 Decennial Census in respondents' Wave IV tract (measured as a G* statistic)
SCST01W5	Census tract residential segregation in 2014-2018 (ACS 5-year) in respondents' Wave V tract (measured as a G* statistic)

Missing codes

Specific missing code information for each variable, including frequency counts, are in the accompanying SCHDSG12 Codebook, and a concise reference table for each variable's missing code is provided at the end of this section.

Missing code formatting logic follows a general "N + 1" standard where N=number of *integer* digits contained in the maximum value for that variable. So, for a variable with maximum legitimate value of 99 (N=2), the missing code format will be 99 + 1 more digit (the number of prefix 9s, or 8s in two variables in this data, always equal to N, with the additional +1 digit suffix specifically identifying the unique code/definition. Missing code prefixes for all variables begin with 9 *except for SCSD01W1 and SCSD01W2* which begin with 8s. Also, please note again that the N for determining missing code digit length is based on total number of *integer* digits of the maximum *valid* value in variable data.

Missing code summary:

- **"Legitimate skip"** 97, 997, 9997, 99997
Used for variables with missing data that logically tracks with missing data from another variable. In the SCHDSG12 data, *"Not Applicable"* is interchangeable with *"Legitimate skip."* For example, if data from Wave I SCSD01W1 (the year a district was dismissed from court order) was not available or showed that the district was never under a court order, then SCSD02W1 (the "First Fall" a district was not under a court order) would show 99997 ("Legitimate skip").

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

- **“Geocode missing”** (98, 998, 9998, 99998)
Indicates a respondent was missing a geocode location in the Add Health source data. This code was used for an Add Health Sample Member (AHSM) participated during the relevant survey wave, but at the time of interview, a valid location geocode was unrecordable/ unavailable at the time.
- **“Missing from source data”** (96, 996, 9996)
Applied when data was missing from one or more external data sources subsequently linked to the AHSM data. In some cases, this was caused by inadequate data being available for proper linkage of the AHSM data.
- **“Missing/No basis for calculation”** (95, 995, 9995, 99995)
Indicates data missing because there was not enough available to properly calculate the values. Where possible, codes for some variables do attempt to distinguish between missing due to lack of geocoding (and therefore, what is incalculable), but in other cases it may have been impossible to entirely distinguish specific reasons for “No basis for calculation” missings.

Additionally, several reasons for an inability to calculate may have existed in the source data beyond no data simply existing in the source. However, code “96” does not provide further details of specific reasons for the inability to calculate, only that it occurred. A few examples might include one or more items from this list:

- (a) Only one tract in the Core-Based-Statistical Area (CBSA)/county
- (b) Zero population in the focal tract
- (c) No target population in the CBSA/county
- (d) Tract is connected to every other tract in the CBSA
- (e) All tracts are islands (no connections)

- **“Not interviewed in Wave #”** (92, 992, 9992, etc.)
Indicates records in which the subject did not participate in a wave or waves.
- **SCSD01W1/SCSD01W2-Year school district was dismissed**
These unique missing codes begin with a four-digit 8888 prefix instead of 9999. The fifth and final digit of code for missing has a specific definition. As described earlier in Section 1 *Data Structure and Form*, the missing codes for these variables describe situations *both* where the data was unknown/missing, *as well as* situations in which a school district was never under a court order or has yet to be released from a court order.
- **SCSD15 If a respondent changed districts between Waves I and II**
Specific unique missing codes (as described in the *Data Structures and Form* sections) that can be used for determining whether a record had data imputed across waves as described in previous sections above. Certain records that were missing relevant geocodes for a particular wave but contained valid ones for others had their location geocode imputed across the waves. (For example, a record with a valid Wave I code but missing a Wave II location was assumed to have the same Wave I location across waves; in that case the Wave II geocode was imputed identically to that in Wave I.)

Detailed information including frequencies of individual variable codes are in the accompanying *SCHDSG12 Codebook*. Also, a general outline of all the missing codes appearing in this dataset along with their definitions are listed below.

Variable Name	Code	Definition
SCSR01W1	97	N/A Legitimate skip
	98	Geocode missing
SCSD01W1	88880	Districts were never under a court desegregation order
	88881	Not dismissed (through 2021)
	88882	Dismissed before 1990 (date uncertain)
	88885	Status and/or dismissal date unknown
	88886	Not dismissed by 2020 (not under order by 1991)
	99998	Geocode missing
SCSD02W1	99996	Missing in the source data
	99997	N/A Legitimate skip
	99998	Geocode missing
SCSD03W1	95	Missing/No basis for calculation
	98	Geocode missing
SCSD04W1	95	Missing/No basis for calculation
	98	Geocode missing
SCSD05W1	95	Missing/No basis for calculation
	98	Geocode missing
SCSD06W1	95	Missing/No basis for calculation
	98	Geocode missing
SCST01W1	995	Missing/No basis for calculation
	998	Geocode missing
SCSD07W1	95	Missing/No basis for calculation
	98	Geocode missing
SCSD08W1	95	Missing/No basis for calculation
	98	Geocode missing
SCSD09W1	9996	Missing in the source data
	9998	Geocode missing
SCSD10W1	96	Missing in the source data
	98	Geocode missing
SCSD11W1	96	Missing in the source data
	98	Geocode missing
SCSD12W1	96	Missing in the source data
	98	Geocode missing
SCSD13W1	96	Missing in the source data
	98	Geocode missing
SCSD14W1	999996	Missing in the source data
	999998	Geocode missing
SCSR01W2	92	Not interviewed at Wave II
	97	N/A Legitimate skip
	98	Geocode missing
SCSD01W2	88880	Districts were never under a court desegregation order
	88881	Not dismissed (through 2021)

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

Variable Name	Code	Definition
	88882	Dismissed before 1990 (date uncertain)
	88885	Status and/or dismissal date unknown
	88886	Not dismissed by 2020 (not under order by 1991)
	99992	Not interviewed at Wave II
	99998	Geocode missing
SCSD02W2	99992	Not interviewed at Wave II
	99996	Missing in source data
	99997	Legitimate skip
	99998	Geocode missing
SCSD03W2	92	Not interviewed at Wave II
	95	Missing/No basis for calculation
	98	Geocode missing
SCSD04W2	92	Not interviewed at Wave II
	95	Missing/No basis for calculation
	98	Geocode missing
SCST01W2	992	Not interviewed at Wave II
	995	Missing/No basis for calculation
	996	Missing in source data
	998	Geocode missing
SCSD07W2	92	Not interviewed at Wave II
	95	Missing/No basis for calculation
	98	Geocode missing
SCSD08W2	92	Not interviewed at Wave II
	95	Missing/No basis for calculation
	98	Geocode missing
SCSD09W2	9992	Not interviewed at Wave II
	9996	Missing in source data
	9998	Geocode missing
SCSD10W2	92	Not interviewed at Wave II
	96	Missing in source data
	98	Geocode missing
SCSD11W2	92	Not interviewed at Wave II
	96	Missing in source data
	98	Geocode missing
SCSD12W2	92	Not interviewed at Wave II
	96	Missing in the source data
	98	Geocode missing
SCSD13W2	92	Not interviewed at Wave II
	96	Missing in source data
	98	Geocode missing
SCSD14W2	999992	Not interviewed at Wave II
	999996	Missing in source data
	999998	Geocode missing
SCSD15	81	Missing Block ID for Wave I
	82	Missing Block ID for Wave II

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

Variable Name	Code	Definition
	84	Source data missing LEAID for Wave II
	85	Source data missing LEAID for Waves I-II
	92	Not interviewed/Not part of the longitudinal sample
SCSD16	92	Not interviewed at Wave II
	95	Missing/No basis for calculation
	98	Geocode missing
SCSD17	92	Not interviewed at Wave II
	95	Missing/No basis for calculation
	98	Geocode missing
SCST01W3	992	Not interviewed at Wave III
	995	No Basis for Calculation
	998	Geocode missing
SCST01W4	992	Not Interviewed at Wave IV
	995	No basis calculation
	998	Geocode missing
SCST01W5	992	Not Interviewed at Wave IV
	995	No basis calculation
	998	Geocode missing

References

1. Reardon SF, Grewal ET, Kalogrides D, Greenberg E. Brown Fades: The End of Court-Ordered School Desegregation and the Resegregation of American Public Schools. *Journal of Policy Analysis and Management*. 2012;31(4):876-904.
2. Johnson RC. *Long-run Impacts of School Desegregation & School Quality on Adult Attainments*. National Bureau of Economic Research, Inc.; 2011. 16664.
3. Lutz B. The End of Court-Ordered Desegregation. *American Economic Journal: Economic Policy*. 2011;3(2):130-168.
4. Liebowitz DD. Ending to What End? The Impact of the Termination of Court-Desegregation Orders on Residential Segregation and School Dropout Rates. *Educational Evaluation and Policy Analysis*. 2018;40(1):103-128.
5. Reardon SF, Grewal ET, Kalogrides D, Greenberg E. District Court Order Data. In. Stanford University: Stanford Center for Education Policy Analysis; 2012. Available at <https://cepa.stanford.edu/data/district-court-order-data>. Accessed April 23, 2021.
6. ProPublica. A national survey of school desegregation orders. ProPublica. <https://projects.propublica.org/graphics/desegregation-orders>. Published 2014. Accessed April 23, 2021.
7. Massey DS, Denton NA. The Dimensions of Residential Segregation. *Social Forces*. 1988;67(2):281-315.
8. Reardon SF, Owens A. 60 Years After Brown: Trends and Consequences of School Segregation. *Annual Review of Sociology*. 2014;40(1):199-218.
9. Getis A, Ord JK. The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*. 1992;24(3):189-206.
10. Ord JK, Getis A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*. 1995;27(4):286-306.

Waves I-V of Add Health were funded by grant P01 HD31921 (Harris) from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.