

2020

Add Health Wave I Documentation



Report prepared by

Rucker Johnson

# Add Health Wave I School District Data Documentation



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.



## Introduction

The SCHDESG1 data file contains measures of average per-pupil spending during school-age years, the average racial school segregation during school-age years, and the number of school-age years of exposure to desegregation court orders or release of them (if applicable) for Add Health respondent residence at Wave I, as measured by U.S. Census block groups mapped to 1969 school district boundaries.

There were two complementary methods used to link the Add-Health data to the school-related variables: (1) ICPSR, (2) GIS. The ICPSR method is based on the “69-70 School District Geographic Reference File” (Bureau of Census, 1970) that relates census tract and school district geographies. One of the disadvantages of this method of merging is that some portions of the US in 1970 did not have a census tract; consequently, this approach fails for locations not assigned a tract. The second method of linking the Add-Health data to school district data is via GIS-based methods using the universe of addresses of schools in 1970 and matching Add-Health respondents to their nearest school. The school district data based on school identification is then obtained. The two different methods of matching respondents to districts (the ICPSR and GIS approach) in combination aims to minimize the total number of missing information that would otherwise arise due to the matching process.

## District-Level K-12 School Spending Data

Education funding data come from several sources that are combined to form a panel of per-pupil spending for US school districts in 1967 and annually from 1970 through 2010. School segregation data come from the Office of Civil Rights (OCR) and Common Core Data (CCD), and are combined to form a panel used to construct school segregation indices that span the period 1968 through 2010. The school segregation and spending data are then linked to a database of desegregation litigation (including the year of initial desegregation court orders and release of court order (if any) between 1954 and 2013.

Data from the Historical Database on Individual Government Finances (INDFIN) combines data from the Census of Governments Survey of Government Finances (F-33), the National Archives, and the Individual Government Finances Survey. The School District Finance Data FY 1967-91 is available annually from 1967 through 1991. It contains over one million individual local government records, including counties, cities, townships, special districts, and independent school districts.

The Common Core of Data (CCD) School District Finance Survey (F-33) consists of data submitted annually to the National Center for Education Statistics (NCES) by state education agencies (SEAs) in the 50 states and the District of Columbia. The purpose of the survey is to supply finance data for all local education agencies (LEAs) that provide free public elementary and secondary education in the United States. Both NCES and the Governments Division of the U.S. Census Bureau collect public school system finance data, and they collaborate in their efforts to gather these data. We combine these data sources to construct per-pupil spending for each school district in the United States for the temporal period of interest. Per-pupil spending data from 1968 and 1969 is missing for all states. Where there was only a year or two of missing per-pupil expenditure data, it was filled in using linear interpolation.

## Racial School Segregation Measures

### Black-White Dissimilarity Index

This index measures the evenness with which black and white students are distributed across schools in the district. Its minimum value is zero and its maximum value is 100.

Suppose:

$b_i$  = the number of black students in the  $i^{\text{th}}$  school

$B$  = the total black student enrollment in the district for which the index of dissimilarity is being calculated.

$w_i$  = the number of white students in the  $i^{\text{th}}$  school

$W$  = the total white student enrollment in the district for which the index of dissimilarity is being calculated

Then, the index of dissimilarity measures the segregation of whites from blacks

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

The summation is over the component schools in the district. The value of this index is statistically independent of the relative size of the groups used in its computation.

### Black-White Exposure Index (measure of school segregation)

$t_i$  = the total student enrollment in the school for which the exposure index is being calculated.

Then, the average percent black in the school of the typical or average white student will be calculated as:

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

The summation is over all the component schools in the district for which the exposure index is being calculated. The maximum value of the black-white exposure index is the percent of black students in the district. That is, if black students make up 30 percent of the students in the district, the maximum value of the average percent black for white students of that district will be 30 percent. This will require that the black-white dissimilarity index measuring the evenness with which blacks and whites are distributed across schools in the district equals zero. The minimum value of the exposure index is zero. That is, although blacks might make up 30 percent of a district, white students could attend exclusively white schools. If the exposure index equals zero, then the dissimilarity index comparing those two racial groups will equal zero. The exposure index involves two mutually exclusive racial groups. However, the average percent black in the school of the typical white student in a district is almost always different from the average percent white in the school for the typical black student in the same district. The formula for calculating the exposure index for any two races may be obtained by making the appropriate changes in notation.

(Source: Rucker C. Johnson, “Long-Run Impacts of School Desegregation and School Quality on Adult Attainments,” NBER Working Paper No. 16664 (revised August 2015), <https://doi.org/10.3386/w16664>.)

## School Desegregation Court Order Measures

The American Communities Project at Brown University compiled a database of every desegregation order, including when it was initially enacted. Desegregation court order timing data was obtained by combining American Communities Project data (Brown University, John Logan) assembled by legal scholars with data from Welch/Light on desegregation plan implementation dates in large districts. See Finis Welch and Audrey Light, *New Evidence on School Desegregation* (Washington, DC: US Commission on Civil Rights, 1987). Desegregation court order timing data obtained by combining American Communities Project data, Welch and Light, *New Evidence*, and data compiled by ProPublica: see Yue Qiu and Nikole Hannah-Jones, “A National Survey of School Desegregation Orders,” ProPublica, December 23, 2014, <https://projects.propublica.org/graphics/desegregation-orders>.

The information on the “year in which a district was released of its desegregation court order” was obtained from original data compiled by Sean Reardon, updated with information from American Communities Project data (Brown University) and ProPublica data on status of desegregation court cases (Yue Qiu and Nikole Hannah-Jones, “A National Survey of School Desegregation Orders”); Sean F. Reardon, Elena Tej Grewal, Demetra Kalogrides, and Erica Greenberg, “Brown Fades: The End of Court-Ordered School Desegregation and the Resegregation of American Public Schools,” *Journal of Policy Analysis and Management* 31, no. 4 (2012): 876–904; Byron Lutz, “The End of Court-Ordered Desegregation,” *American Economic Journal: Economic Policy* 3, no. 2 (2011): 130–168, <https://doi.org/10.1257/pol.3.2.130>.

The combined data from the American Communities Project (Brown University), Welch/Light, Sean Reardon, and Propublica, provide the best available data that have ever been utilized to study this topic. Based on the child residential location at the Wave I survey, mapped onto the school district boundaries that prevailed in 1969, the average per-pupil spending during school-age years, the average racial school segregation during school-age years, and the number of school-age years of exposure to desegregation court orders or release of them (if applicable) are computed, respectively.

(Source: Rucker C. Johnson, “Long-Run Impacts of School Desegregation and School Quality on Adult Attainments,” NBER Working Paper No. 16664 (revised August 2015), <https://doi.org/10.3386/w16664>.)

## Variable Nomenclature

The variables in the data that utilized the ICPSR method of linking have a suffix of “I” in their names. Similarly, variables linked using the GIS method are denoted with a suffix of “G” in their names.

## Missing data

Missing information for desegregation court order is simply an artifact of court orders only occurring in a subset of U.S. school districts; nonetheless, there exists a large degree of overlap between districts that were subject to desegregation court orders (or had them lifted) and places in which Add Health respondents lived as children. Moreover, missing information for the desegregation court order variables typically represents a district that was never subject to court order. Missing values for a relatively small portion of the school spending measures results from the consolidation of districts in some instances and the fact that the data rely on 1969 school district boundaries in order to address endogeneity concerns.

In the Wave I School District Data, two special codes distinguish types of missing data. They are as follows:

Value	Reason for missing data
99997	School District was either not dismissed from desegregation or never under a court order to desegregate
98, 9998, 99998, 99999998	Geocode missing at Wave I

## Data Dictionary

Name	Description
SDSEG01	Average per-pupil spending during ages 5-17 in district of upbringing (expressed in real 2012 dollars)
SDSEG02	School district enrollment size in 1976
SDSEG03D	Average racial school segregation during ages 5-17 in district of upbringing (as measured by black-white dissimilarity index)
SDSEG03E	Average racial school segregation during ages 5-17 in district of upbringing (as measured by black-white exposure index)
SDSEG04I	Year of initial desegregation court order in district of upbringing (if any) (using ICPSR-method of matching)
SDSEG04G	Year of initial desegregation court order in district of upbringing (if any) (using GIS-method of matching)
SDSEG05I	Year of release of desegregation court order in district of upbringing (if any) (using ICPSR-method of matching)
SDSEG05G	Year of release of desegregation court order in district of upbringing (if any) (using GIS-method of matching)
SDSEG06I	<p>Categorical variables of the legal status of the desegregation court order (ICPSR -method of matching, respectively):</p> <p>Labels 0=“Don’t have status”  1=“Dismissed, have date”  2=“Dismissed, no date”  3=“Not dismissed”  4=“Never under court order”</p>
SDSEG06G	<p>Categorical variables of the legal status of the desegregation court order (GIS-method of matching, respectively):</p> <p>Labels 0=“Don’t have status”  1=“Dismissed, have date”  2=“Dismissed, no date”  3=“Not dismissed”  4=“Never under court order”</p>
SDSEG07	This is a categorical grouping variable to be used for the inclusion of school district fixed effects, and/or to cluster standard errors at the district level, in regression models.
SDSEG08	This is a categorical grouping variable to be used for the inclusion of childhood county fixed effects, and/or to cluster standard errors at the county level, in regression models.