

ICPSR 21600

**National Longitudinal Study of  
Adolescent to Adult Health (Add  
Health), 1994-2008 [Public Use]**

Kathleen Mullan Harris  
*University of North Carolina-Chapel Hill*

J. Richard Udry  
*University of North Carolina-Chapel Hill*

Data Guide

Inter-university Consortium for  
Political and Social Research  
P.O. Box 1248  
Ann Arbor, Michigan 48106  
[www.icpsr.umich.edu](http://www.icpsr.umich.edu)

## Terms of Use

The terms of use for this study can be found at:  
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/21600/terms>

## Information about Copyrighted Content

Some instruments administered as part of this study may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

### NOTICE

#### WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

# National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008 [Public Use] (ICPSR 21600)

## Table of Contents

- I. Introduction**
- II. Sample**
- III. Data Elements**
- IV. Variables**
- V. Weights**
- VI. Merging Data Files**
- VII. How to Obtain Data and Documentation Files**
- VIII. Learn More**

## I. Introduction

### About the Guide

This Data Guide is an overview of the National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008 [Public Use] (ICPSR 21600) and provides specific instructions for obtaining the Add Health datasets, which you can download to your own computer from [DSDR](#) at ICPSR. The Data Guide is also available as a [webpage](#). Add Health users should also refer to the [User Guide](#) (pdf), which provides greater detail on the topics discussed below.

### About the Data

The National Longitudinal Study of Adolescent to Adult Health (Add Health) was developed in response to a mandate from the U.S. Congress to fund a study of adolescent health. Initiated in 1994, Add Health has been supported by four program project grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH) with co-funding from 23 other federal agencies and foundations. Designed by researchers at the University of North Carolina, Add Health is the largest, most comprehensive longitudinal survey of adolescents ever undertaken. Beginning with an in-school questionnaire administered to a nationally representative sample of students in grades 7-12 during the 1994-95 school year, the study followed up with a series of in-home interviews conducted in 1995, 1996, 2001-02, 2008, and 2016-2018<sup>1</sup>. Other sources of data include questionnaires for parents, siblings, fellow students and school administrators, and interviews with romantic partners<sup>2</sup>. Preexisting databases provide information about neighborhoods and communities.

---

<sup>1</sup> Add Health is still in the field doing data collection.

<sup>2</sup> Siblings, romantic partners, and school administrative data are only available with a restricted-use contract.

Add Health consists of five waves of data. Each wave combines longitudinal survey data on respondents' social, economic, psychological and physical well-being with contextual data on the family, neighborhood, community, school, friendships, peer groups, and romantic relationships, providing unique opportunities to study how social environments and behaviors in adolescence are linked to health and achievement outcomes in young adulthood. Multiple datasets are available for study from each wave of data, providing opportunities to increase knowledge in the social and behavioral sciences and many theoretical backgrounds. A brief summary of each wave follows:

**Waves I and II**, conducted in 1994-95 and 1996 respectively, focus on the forces that may influence adolescents' health and risk behaviors, including personal traits, families, friendships, romantic relationships, peer groups, schools, neighborhoods, and communities. As participants have aged into adulthood, however, the scientific goals of the study have expanded and evolved.

**Wave III**, conducted in 2001-02 when respondents were between 18 and 26<sup>3</sup> years old, focuses on how adolescent experiences and behaviors are related to decisions, behavior, and health outcomes in the transition to adulthood.

**Wave IV**, conducted in 2008, respondents were ages 24-32<sup>4</sup> and assuming adult roles and responsibilities. Follow up at Wave IV has enabled researchers to study developmental and health trajectories across the life course of adolescence into adulthood using an integrative approach that combines the social, behavioral, and biomedical sciences in its research objectives, design, data collection, and analysis. The fourth wave of interviews expanded the collection of biological data in Add Health to understand the social, behavioral, and biological linkages in health trajectories as the Add Health cohort ages through adulthood

**Wave V**<sup>5</sup> data collection began in 2016 and continues the biological data expansion that began in Wave IV. During the period of 2016-2018, social, environmental, behavioral, and biological data will be collected to track the emergence of chronic disease as the cohort moves through their fourth decade of life.

---

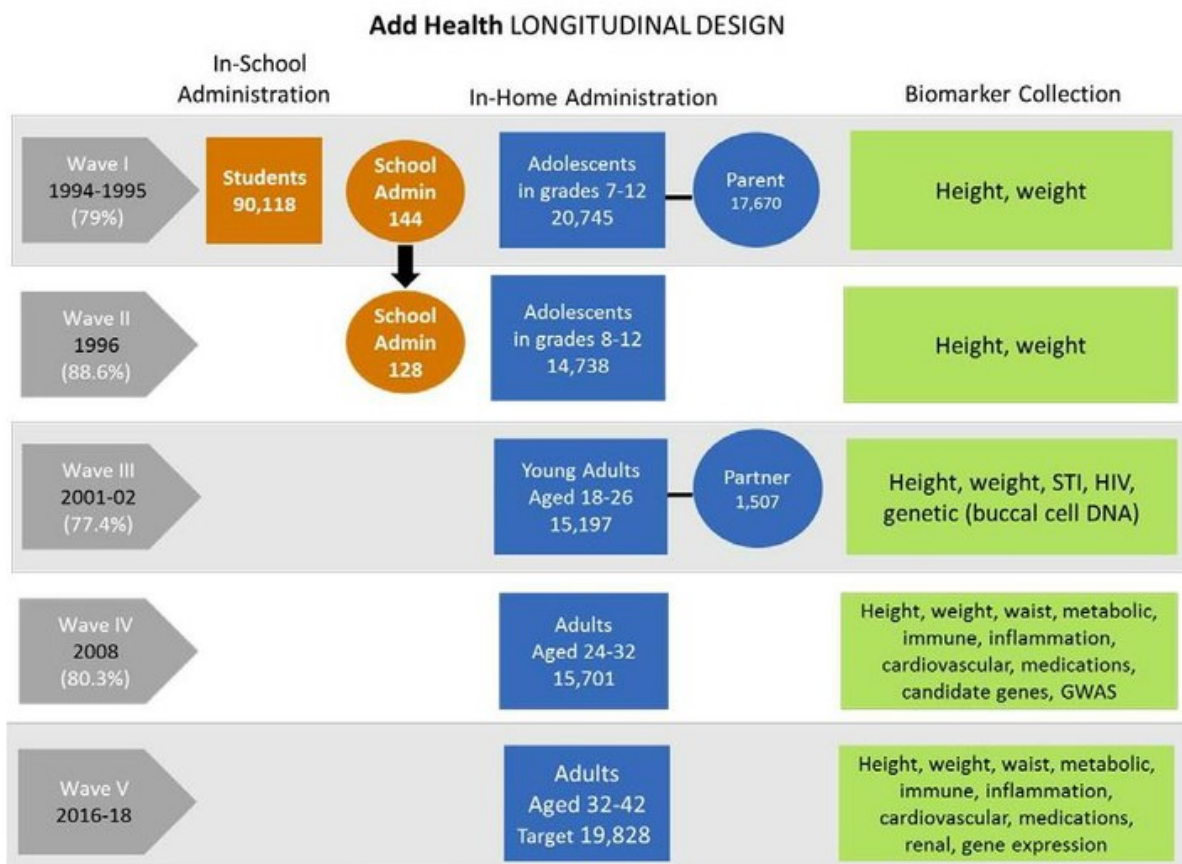
<sup>3</sup> 24 respondents were 27-28 years old at the time of the Wave III interview.

<sup>4</sup> 52 respondents were 33-34 years old at the time of the Wave IV interview.

<sup>5</sup> Wave V is not currently available for secondary analysis; Wave V Sample 1 is only available as a restricted-use file.

## II. Sample

The following chart depicts the sampling structure for Add Health:



Detailed information on the sampling for each wave can be found on the [Add Health website](#). The [Add Health Research Design presentation](#) (pdf) provides additional information about study design for Waves I-V. Please note the public-use dataset consists of one-half of the core sample, and one-half of the oversample of African-American adolescents with a parent who has a college degree, chosen at random. This is roughly 1/3 of the full sample. N's will not match between the restricted-use and public-use data.

Unless appropriate adjustments are made for sample selection and participation, estimates from analyses using the Add Health data can be biased when any factor used as a basis for selection as a participant in the Add Health Study also influences the outcome of interest. For example, black adolescents whose parents were college graduates comprise one of the many over-sampled groups. Parental education is a factor that affected selection of black youth in the Add Health study and can also influence family income. Unless the analytic technique uses appropriate statistical methods to adjust for over-sampling, estimates of the income of blacks will be biased. Any analysis that includes family income, or other variables related to family income, may produce biased estimates unless proper adjustments are made for over-sampling.

To obtain unbiased estimates, it is important to account for the sampling design by using analytical methods designed to handle clustered data collected from respondents with unequal probability of selection. Failure to account for the sampling design usually leads to under-estimating standard errors and false-positive statistical test results. Please see the [User Guide](#) (pdf) for a list of the attributes of the Add Health sampling design that should be taken into consideration during analysis.

### III. Data Elements

The Add Health data are available in two forms—public-use (files listed in Table 1) and restricted-use. It is a central concern of the Add Health study that the confidentiality of respondents be strictly protected. Deductive disclosure concerns prevent full access to all data sources. The [restricted-use dataset codebooks](#) and an online interactive [Add Health Codebook Explorer \(ACE\)](#) are available for further exploration. Wave V Sample 1 is only available as a restricted-use file. The complete Wave V data will be available in both public-use and restricted-use formats. To apply for a restricted-use dataset, please see the [Add Health Contracts page](#).

Public-use data for Add Health are collected from multiple sources and made available in 31 data sets (see Table 1 below). Data are available for study from three instruments in Wave I (conducted from September 1994 through December 1995), one survey in Wave II (conducted from April 1996 through August 1996), several sources in Wave III (collected from August 2001 through April 2002), and one in-home interview in Wave IV (conducted from January 2008 through February 2009).

**Table 1. List of Available Public-Use Data Files**

Part Number	File Name
DS1	Wave I: In-Home Questionnaire, Public Use Sample
DS2	Wave I: Public Use Contextual Database
DS3	Wave I: Network Variables
DS4	Wave I: Public Use Grand Sample Weights
DS5	Wave II: In-Home Questionnaire, Public Use Sample
DS6	Wave II: Public Use Contextual Database
DS7	Wave II: Public Use Grand Sample Weights
DS8	Wave III: In-Home Questionnaire, Public Use Sample
DS9	Wave III: In-Home Questionnaire, Public Use Sample (Section 17: Relationships)
DS10	Wave III: In-Home Questionnaire, Public Use Sample (Section 18: Pregnancies)
DS11	Wave III: In-Home Questionnaire, Public Use Sample (Section 19: Relationships in Detail)
DS12	Wave III: In-Home Questionnaire, Public Use Sample (Section 22: Completed Pregnancies)
DS13	Wave III: In-Home Questionnaire, Public Use Sample (Section 23: Current Pregnancies)
DS14	Wave III: In-Home Questionnaire, Public Use Sample (Section 24: Live Births)
DS15	Wave III: In-Home Questionnaire, Public Use Sample (Section 25: Children and Parenting)
DS16	Wave III: Public Use Education Data
DS17	Wave III: Public Use Graduation Data
DS18	Wave III: Public Use Education Data Weights
DS19	Wave III: Add Health School Weights
DS20	Wave III: Peabody Picture Vocabulary Test (PVT), Public Use

DS21	Wave III: Public In-Home Weights
DS22	Wave IV: In-Home Questionnaire, Public Use Sample
DS23	Wave IV: In-Home Questionnaire, Public Use Sample (Section 16B: Relationships)
DS24	Wave IV: In-Home Questionnaire, Public Use Sample (Section 16C: Relationships)
DS25	Wave IV: In-Home Questionnaire, Public Use Sample (Section 18: Pregnancy Table)
DS26	Wave IV: In-Home Questionnaire, Public Use Sample (Section 19: Live Births)
DS27	Wave IV: In-Home Questionnaire, Public Use Sample (Section 20A: Children and Parenting)
DS28	Wave IV: Biomarkers, Measures of Inflammation and Immune Function
DS29	Wave IV: Biomarkers, Measures of Glucose Homeostasis
DS30	Wave IV: Biomarkers, Lipids
DS31	Wave IV: Public Use Weights

## IV. Variables

Variable names are constructed to provide information regarding data collection method, wave of data collection, interview section title, and question number. Typically, the first two alphanumeric characters in the variable name indicate the data collection method (H=in-home interview, S=in-school questionnaire, and P=parent questionnaire) and wave (1-4) of data collection. The next two alphanumeric characters are an abbreviation for the interview section title. The question number is in the last 5 to 8 alphanumeric characters. However, there are exceptions. Non-interview data variables are usually mnemonic, such as SMPxx for sample. Lab results are often abbreviations for the test performed. Constructed variables usually begin with a C followed by a mnemonic or number.

Add Health restricted-use data variables from Wave I-IV can be explored through the [Add Health Codebook Explorer \(ACE\)](#). The ACE can be browsed by topic or search for questions by variable name, keyword, or phrase in order to discover the rich volume of data collected by Add Health. The questions are organized by topic, subtopic, and variable. Collections of variables were constructed specifically for this site to show similar questions asked across several waves of data collection and do not represent grouping for research purposes. Variables in Add Health can also be searched and compared directly from the [DSDR Add Health study home page](#).

## V. Weights

The Add Health sampling weights are designed to turn the sample of adolescents interviewed into the population desired for study. These weights are available for the respondents who are members of the Add Health probability sample. By using these sampling weights and a variable to identify clustering of adolescents within schools, unbiased estimates of population parameters and standard errors can be obtained from analysis. Please see [Chapter 2 of the User Guide](#) (pdf) for descriptions and detailed tables of the sampling weights distributed with the Add Health data and instructions on which weight should be used in analysis. The Add Health sampling weights were developed for analyzing combinations of data from the In-Home Interviews using a variety of techniques. Usage of these weights can be divided into three different categories of analyses:

Single-Level (Population-Average) Model, Multilevel Model, and Single-Level Model for Special Subpopulations. The sampling weight selected for an analysis depends on both the type of analysis required to investigate a hypothesis and the interview or combination of interviews needed in the analysis. Weights are given for Cross-Sectional Analysis, Longitudinal Analysis, and Time-to-Event Analysis. The guidelines presented in Chapter 2 for choosing the correct sampling weight for most analyses can be summarized in three simple rules:

1. Cross-Sectional Analysis: Choose the weight created for everyone in the probability sample (see User Guide, Table 2.4) for the population of interest.
2. Longitudinal Analysis: Choose the weight from the Wave of data collected at the latest time-point (see User Guide, Table 2.5) for the population of interest.
3. Time-to-Event Analysis: Choose the weight from the Wave of data collected at the earliest time point (see User Guide, Table 2.6) for the population of interest.

These rules should allow the analyst to select the best sampling weight for most research endeavors. Additional information on the longitudinal weights can be found in the [Wave I, III, & IV Longitudinal Weight for Public-Use Sample User Guide](#) (pdf).

The [User Guide](#) (pdf) discusses how to correct for design effects and the unequal probability of selection to ensure that analysis results are nationally representative with unbiased estimates, but it refers to variables from the Add Health Restricted-Use Data. For the public-use data, CLUSTER2 should be used in conjunction with the correct weight variable. A table describing this is available from [Add Health's FAQ page](#).

The User Guide also includes a chapter on avoiding common errors that occur when analyzing Add Health data. Common errors addressed in this section are:

- Ignoring clustering and unequal probability of selection when analyzing the Add Health data
- Including respondents who are missing sampling weights in analyses when your goal is to obtain national estimates
- Subsetting the probability sample (i.e., adolescents who have weights) when using the survey software
- Using the Sampling Weight as a Frequency or Analytical Weight during Analysis
- Normalizing the Sampling Weights

Please see [Chapter 3](#) (pdf) for more details regarding steps to prepare the data for analysis.

## VI. Merging Data Files

The public-use datasets should be merged using the variable AID. Public-use data doesn't contain ID numbers of friends, siblings or romantic partners, so the data cannot be linked. The following is skeleton SAS and Stata code for merging the data:

### SAS Code

```
/* sorts the input data files */
```



```
proc sort data = < data file1>;
```

```
  by AID;
```

```
run;
```

```
proc sort data = < data file2>;
```

```
  by AID;
```

```
run;
```

```
/* merges the input data files and keeps only the variables of interest */
```

```
data <new file> (keep=var1 var2 var3 var4);
```

```
  merge <data file1> <data file2>;
```

```
  by AID;
```

```
run;
```

## **Stata Code**

```
use "wave1"
```

```
merge 1:1 aid using "wave2"
```

```
merge 1:1 aid using "wave3"
```

```
merge 1:1 aid using "wave4"
```

Note if the original data files are xpt format, use “fdause” and “save” to transform the data sets into stata format first.

## **VII. How to Obtain Data and Documentation Files**

### **Downloading Data and Documentation from DSDR**

Public-use data from the National Longitudinal Study of Adolescent to Adult Health (Add Health), 1994-2008 are made available through [DSDR](#), a data archive within ICPSR.

Researchers interested in downloading analysis-ready data and documentation files can do so free of charge through the [DSDR website](#). Data are available in four statistical package formats: SAS, SPSS, STATA, and R. Raw ASCII and Excel/TSV data are also provided with accompanying setup (syntax) files. Documentation is provided in PDF format.

To download the Add Health data and/or documentation, researchers must agree to the Terms of Use. To download all public-use files, select the [Download drop-down menu](#) from the top left-hand side of the study homepage. Choose the file format you would like.

### **First Steps toward Obtaining Your Analytic Files**

Before downloading the data or beginning analysis, it is important for the user to become familiar with the [Add Health User Guide](#) (pdf).

## VIII. Learn More

### Additional Resources

- [Add Health Website](#)
  - [Add Health FAQ](#)
- [Add Health Publications Database](#)
- [ICPSR Abbreviated bibliography of Add Health-based Publications](#)
- List Servers:
  - Add Health: to interact with other data users and analysts: Send email to [listserv@unc.edu](mailto:listserv@unc.edu) and in the body of the message type: subscribe addhealth2 firstname lastname
  - Add Health: to receive notifications about data and documentation: Send email to [addhealth@unc.edu](mailto:addhealth@unc.edu) and in the subject line put Add Health List Server.
- [Add Health Research Twitter account](#)
- Questions regarding the Add Health project? [Email Add Health](#).
- Questions regarding this Data Guide or Add Health data files? [Email DSDR](#).

### Acknowledgements

This Data Guide was prepared by Sara C. Britt using Add Health documentation created by Add Health project staff from the University of North Carolina at Chapel Hill's Carolina Population Center and ICPSR. It was developed for the Data Sharing for Demographic Research (DSDR), a project supported by the Population Dynamics Branch (PDB) of the Eunice Kennedy Shriver National Institute of Child Health and Human Development (U24 HD048404). DSDR is housed within the Inter-university Consortium for Political and Social Research (ICPSR).