DSDR | DATA SHARING FOR DEMOGRAPHIC RESEARCH

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

Wave II: Public Use Contextual Database Codebook/Questionnaire

Inter-university Consortium for Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
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//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.

ICPSR 21600

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

Wave II: Public Use Contextual Database

Original P.I. Documentation

National Longitudinal Study of Adolescent Health

Public Use Contextual Database

Waves I and II

❖ ❖ ❖ ❖ ❖

John O.G. Billy
Audra T. Wenzlow
William R. Grady

Carolina Population Center
University of North Carolina at Chapel Hill

March 1998

This research was supported by grant P01-HD31921 from the National Institute of Child Health and Human Development. Further information may be obtained by contacting Jo Jones, PhD, Project Manager, 919/962-8412 (email: jo_jones@unc.edu) at the Carolina Population Center, CB# 8120 University Square, Chapel Hill, NC 27516-3997.

User Documentation

for

The Add Health Public Use Contextual Database

by

John O.G. Billy
Audra T. Wenzlow
William R. Grady

Subcontract No. UNC-CH #5-52421
Prime Contract No. 1 P01 HD31921

BATTELLE
Centers for Public Health Research and Evaluation
4000 N.E. 41st Street
Seattle, WA 98105
(206) 525-3130

Carolina Population Center
University of North Carolina at Chapel Hill
123 W. Franklin Street
Chapel Hill, NC 27516-3997

Table of Contents

I. Introduction

There is a growing recognition that the characteristics of the places in which young people live shape their health-related decisions and behaviors by influencing both the alternatives available to the adolescent, and their associated social, economic and psychic costs. The purpose of the public use version of the Add Health (National Longitudinal Study of Adolescent Health) Contextual Database is to provide an array of community characteristics by which researchers may investigate the nature of such contextual influences for a wide range of adolescent health behaviors. Selected contextual variables have been calculated, compiled and are provided here, already linked to the Add Health respondent IDs.

For most respondents participating in the Add Health in-home survey, Wave I and Wave II home locations were identified. When possible, these locations have been geocoded in order to link them to their block group census areas.[1] The availability of block group level data in the 1990 Census of Population and Housing for each of these areas has allowed the creation of two contextual data files corresponding to the two waves of data collection in the Add Health in-home survey. Missing data associated with the geocoding process are described under Missing Data.

The variables contained in the Add Health Public Use Contextual Database are detailed in the Data Dictionary. However, to successfully and accurately access and use data from the database, it is necessary to understand the form of the data, constructed measure characteristics, and the types of missing data values that exist in the files. With this information, specific measures can be accurately identified, and subsequent analyses of these data can be interpreted meaningfully. It is suggested that the remaining sections of the Introduction be read carefully before any contextual data are used.

Documentation Structure

This Introduction to the User Documentation for the Add Health Public Use Contextual Database provides information required to understand the contextual database contents and form and conventions used. The source of the data used to construct measures, the Census of Population and Housing, 1990: Summary Tape File 3A (STF 3A), is described in Source Information. The section entitled Data Form describes the technical structure of the data files. The Constructed Measures section contains a general discussion about the variables included in the database. Variable names are described under Variable Naming Conventions. Finally, types of missing data are detailed under Missing Data.

Following the Introduction, the other main section of this documentation is the Data Dictionary. The Data Dictionary lists each variable with its complete name, and includes references to the appendices. Ordered by subject and listed under subject headings, the Data Dictionary is used to identify variables of interest. The technical appendix, Appendix A, contains definitions for the statistical measures that are used in the construction of the contextual variables. Detailed variable category information, when relevant, is provided in Appendix B. Finally, Appendix C contains a Codebook for each of the two files that comprise the Add Health Public Use Contextual Database. Summary statistics and missing data frequencies are listed in the order that variables reside in each data file.

Source Information

The block group is a U.S. Bureau of the Census defined geographic area, which in 1990, averaged 452 housing units, or 1,100 people.[2] It is the lowest level of geography for which the Census Bureau publishes sample data, and thus captures the most localized available contextual characteristics of the areas in which individuals live. Block group level data from the Census of Population and Housing, 1990: Summary Tape File 3A (STF 3A) have been used to create constructed measures in the Add Health Public Use Contextual Database. The STF 3A is the principal, national-in-scope source of contextual data at the block group level.

The STF 3A contains detailed tabulations of population and housing characteristics produced from the 1990 Decennial Census. It contains over 2,300 "cells" (variables), providing information on age, race, ethnic (Hispanic), and sex composition; marital status; migration; year moved into residence; education; labor force participation; unemployment; income in 1989; poverty status; occupation; household type; etc. These variables are derived from the 1990 Census long-form questionnaire received by approximately one in six housing units in the U.S. Thus, data in the STF 3A are sample data that have been weighted by the Census Bureau to represent the total population of the geographic units to which they pertain.

In the STF 3A, identical data items are available for states and their subareas in hierarchical sequence, from counties down to the census tract/block numbering areas (BNAs) and the block group. In defining and understanding *block group* as a concept, it is first necessary to define other Census geography, especially census tracts and block numbering areas (BNAs). A census tract is "a small locally defined statistical area within selected counties, generally having stable boundaries and, when first established by local committees, designed to have relatively homogeneous demographic characteristics. Census tracts do not cross county boundaries."[3] They are generally defined for metropolitan areas and other highly populated counties and usually contain between 2,500 and 8,000 people.[2] A block numbering area (BNA) is "an area delineated cooperatively by a State and the Census Bureau for grouping and numbering blocks in [all] areas where census tracts have not been established. Block numbering areas do not cross county boundaries."[3] The Census Bureau publishes the same types of data for BNAs as it does for census tracts, thereby treating them as equivalent. In sparsely populated counties, however, the average population size of BNAs will be smaller than that of census tracts.[3]

During the 1990 Census, for the first time, all areas in the U.S. were block-numbered. A census block is "a small, usually compact area, bounded by streets and other prominent physical features as well as certain legal boundaries. Blocks do not cross BNA, census tract, or county boundaries."[3] A block group is a cluster of such blocks, always falling within a tract or BNA. A typical census tract contains four or five block groups.

For more information regarding the STF 3A and Census geographic area delineations, the reader is directed to the technical documentation of the data source.[4]

Data Form

Both contextual data files contain one observation for each respondent in the corresponding wave of the Add Health in-home survey. In the complete Wave I data set, 20,745 respondents were identified in 4,411 different block groups. In Wave II, 14,738 followup interviews were conducted with respondents residing in 3,648 different block groups. The number of block groups may be smaller for the public use samples. Respondents living in the same geographic location at both times have the same contextual data in both files. The contextual data differ for Wave I and Wave II residences only for respondents identified to have moved to a different block group between survey waves based on geocode information.

There are 32 variables in each contextual database file. The first three variables on each file are: an eight character Respondent ID (the **AID**); a **MATCH** variable indicating how the respondent's block group was identified (0 = not matched; 1 = GPS reading;[5] 2 = address match; and 3 = ZIP+4 centroid match in urban area); and a **MOVER** variable indicating whether the respondent changed residential locations between survey waves (0 = not a mover; 1 = moved to a different block group; 2 = moved to a different residence within the same block group; 3 = moved, location of Wave I or Wave II residence unknown; and 9 = respondent did not participate in both waves). These variables are followed by 29 contextual variables describing the characteristics of the block groups within which the respondents reside. Except for the **AID**, each variable is numeric and has the default SAS length of 8 bytes. Variables in the Add Health Public Use Contextual Database files are ordered as they appear in the Data Dictionary, with the Respondent ID, **MATCH**, and **MOVER** variables preceding the contextual variables. The Respondent ID is required for linking the database to Add Health respondent data.

Missing data values for numeric variables are coded as "8"or "9" to distinguish different types of missing data. This method was used to denote missing data due to: (1) those set to missing due to estimability difficulties ("8"); and (2) missing geocodes ("9") (see Missing Data for details). The value labels for these variables are "8," unstable estimates and "9," geocode missing.

Labels have been associated with each variable contained in the Wave I and Wave II contextual data files. However, labels are limited to 40 characters, requiring abbreviated descriptions. Please refer to the Data Dictionary for a complete description of any variable.

Constructed Measures

Constructed measures were chosen for inclusion in the public use version of the Add Health Contextual Database to capture a wide range of contextual influences. Modes and medians were calculated to reflect central tendency for categorical and continuous measures, respectively. Median age is rounded to the nearest whole number; income and housing value medians are rounded to the nearest thousand. Dispersion measures, rounded to the third decimal place, are provided to describe the degree to which characteristics vary within block groups. Several categorical variables are also included to capture demographic characteristics such as the sex composition and female labor force participation within the geographic area. Distributional characteristics and the substantive meaning of each measure were considered in making category determinations. These are detailed in Appendix B.

In a very few block groups, there are two modes for a particular measure. In these cases, of the two modal categories, the one dominant over all block groups was selected to represent the mode. The dispersion measures will identify the near equal distribution of categories within such block groups.

Variable Naming Conventions

Each of the 29 contextual variables is assigned a name beginning with BST90P followed by two unique digits (e.g., **BST90P28**). The BST90P prefix indicates that the variable contains block group data from the STF 3A for 1990, and is part of the public use contextual database.

Note that the last two characters of the variable name refer to a designated variable number within the Add Health Public Use Contextual Database. Variable numbers range from 1 to 29 in both contextual files. Variables are ordered sequentially by the variable number. This sequential order is also the order in which variables are listed in the Data Dictionary.

Missing Data

Two types of missing data have been coded in the contextual data files: missing geocode data and data set to missing due to small sample sizes. Missing data due to unavailable geocodes are set to "9" (preceded by additional 9s for variables longer than one). A value of "8" (preceded by additional 9s for variables longer than one) is used to denote observations set to missing when small sample sizes create estimability problems. Missing data were excluded hierarchically according to this order. Both types of missing data are described below in more detail.

Most respondent addresses were accurately matched to identify the block group of the Wave I and Wave II residence. Further, for the majority of those addresses that could not be matched, GPS readings were taken that allow the accurate geocoding of residence locations. For some respondents, however, information about residential location at one or both survey waves was limited to ZIP code data. That is, addresses that could not be matched and for which GPS readings were not available were associated with a census location based on the centroid of the residence ZIP+4 code. Types of matches can be determined by the **MATCH** variable that is included in each data file. If an address was not matched, all variable values for that observation are set to missing ("9"). In addition, ZIP+4 matched addresses in rural areas were set to missing. Table 1

summarizes geocode missing data determinations, showing the frequency of each match type by survey wave.

Table 1. Type of Geocode Match by Survey Wave for Public Use Contextual Database

Level of Match	**Wave I**	**Wave II**
GPS Reading	1,869	1,408
Address Match	4,451	3,279
ZIP+4 Match/Urban	111	84
No Match	73	66

In the Add Health Public Use Contextual Database, the estimates of the characteristics of the geographic units are based on sample data rather than data from the population as a whole. Specifically, these are estimates based on STF 3A data that were derived from the Census long-form questionnaire administered to only about one in six or seven housing unit residents. Because these are estimates based on sample data, confidence that they reflect true population values declines as the size of the sample on which they are based declines. For this reason, estimates were set to missing ("8") when there was evidence that they were unacceptably unstable due to very small sample sizes.

In general, two different standards for determining what is a sufficient sample size were adopted. For estimates based on the dichotomous responses of individuals (e.g., proportions at the aggregate level that are based on yes/no responses at the individual level), estimates were set to a missing value when the estimated population size of the aggregate unit was smaller than 70, indicating a sample size of less than 10. For estimates based on continuous variables (e.g.,income) or variables with a large number of response categories (e.g., occupation) estimates were set to a missing value when the estimated population size of the aggregate unit was smaller than 170, indicating a sample size of less than 25. These different criteria were used because variables of the latter type have larger variances and require larger samples to produce stable estimates. Note that the number of people residing in some block groups that were newly developed between the time of the 1990 Census and the Add Health in-home survey may have been very small or even zero at the time of Census enumeration. Constructed measures based on these zero or very small counts will contain missing data according to the sample size criteria noted above.

Frequencies of special missing value codes in the Add Health Public Use Contextual Database are included in Appendix C - Contextual Database Codebook for each variable in the Wave I and Wave II data files. None of the constructed measures has been deleted from these data files on the basis of the number of missing cases.

Notes

1. Geocodes for the Wave I and Wave II home locations were provided to Battelle by the Carolina

Population Center at the University of North Carolina at Chapel Hill (UNC-CH) in conjunction with the National Opinion Research Center (NORC), the contractor responsible for conducting the fieldwork of the Add Health Study.

2. Bureau of the Census. 1993. *A Guide to State and Local Census Geography*. Publication 1990 CPJ-I-18. Washington, DC: U.S. Government Printing Office.

3. Bureau of the Census. 1990. *TIGER: The Coast to Coast Digital Map Data Base*, p. 17. Washington, DC: Data User Services Division.

4. Bureau of the Census. 1992. *Census of Population and Housing, 1990: Summary Tape File 3 on CD-ROM Technical Documentation*. Washington, DC: Bureau of the Census.

5. Global Positioning System reading of the longitude and latitude coordinates of the adolescent's home.

II. Data Dictionary

The Data Dictionary lists all the variables contained in the two data files that comprise the Add Health Public Use Contextual Database. Preceding the contextual variables are three linking/geocoding/mover codes that are appended to each of the data files, and described in the Introduction. All contextual variables included in the database are of numeric type and are listed in the Data Dictionary by subject, beginning with population measures and concluding with housing characteristics. The user is referred to Appendix A for further definition of statistical measures, and Appendix B for variable category details.

Data Dictionary - Linking/Geocoding Variable List

Variable Name	*Description*
AID	Respondent ID
MATCH	Geocode match indicator 0 = no match 1 = GPS match 2 = address match 3 = ZIP+4 match/urban
MOVER	Mover indicator: respondent moved between Wave I and Wave II 0 = respondent did not move 1 = moved to different block group 2 = moved within same block group 3 = moved, location unknown 9 = respondent did not participate in both waves

Data Dictionary - Contextual Variable List

Variable Name	*Description*
	POPULATION
	Distribution
BST90P01	Urbanicity code[1] 1=completely urban 2=not completely urban
	Race, Sex, and Age
BST90P02	Modal race 1=white 2=black 3=other
BST90P03	Dispersion in race composition (white/black/other)
BST90P04	Proportion Hispanic[1] 1=low 2=medium 3=high 4=very high
BST90P05	Sex composition[1] 1=heavily male 2=balanced 3=heavily female

Variable Name	*Description*
BST90P06	Median age (10 year age categories, and 80+)
BST90P07	Dispersion in age distribution (10 year age categories, and 80+)

VITAL STATISTICS

Marital Status

BST90P08	Modal marital status (excludes persons not in these categories) 1=never married 2=married, spouse present 3=separated or divorced
BST90P09	Dispersion in marital status (never married/married, spouse present/separated or divorced)

Fertility Indicator

BST90P10	Proportion population that are children under five years old[1] 1=low 2=medium 3=high

Migration

BST90P11	Modal migration status 1=lived in same house in 1985 2=lived in different house in 1985, same county 3=lived in different house in 1985, different county
BST90P12	Dispersion in migration status (lived in same house in 1985/lived in different house in 1995, same county/lived in different house in 1985, different county)

HOUSEHOLDS

BST90P13	Modal household type 1=married couple family household 2=other family household 3=non-family household
BST90P14	Dispersion in household type (married couple family household/other family household/non-family household)

INCOME

Household Income in 1989

BST90P15	Median household income (9 income categories)[1]
BST90P16	Dispersion in household income distribution (9 income categories)[1]

Family Income in 1989

BST90P17	Median family income (9 income categories)[1]
BST90P18	Dispersion in family income distribution (9 income categories)[1]

POVERTY STATUS

BST90P19 Proportion persons with income in 1989 below poverty level (for persons for whom poverty status is determined)[1]

1=low
2=medium
3=high

EDUCATION

BST90P20 Modal educational attainment of individuals aged 25 years and over

1=no high school degree or equivalency
2=high school degree, no college degree
3=college degree or more

BST90P21 Dispersion in educational attainment of individuals aged 25 years and over (no high school degree or equivalency/high school degree, no college degree/college degree or more)

LABOR FORCE

Female Labor Force Participation

BST90P22 Proportion females aged 16 years and over in the civilian labor force[1]

1=low
2=medium
3=high

Unemployment

BST90P23 Unemployment rate[1]

1=low
2=medium
3=high

Occupation

BST90P24 Modal occupation type for employed persons 16 years and over

1=managerial or professional
2=technical, sales or administrative support
3=service occupations
4=farming, forestry or fishing
5=production, craft or repair
6=operators, fabricators and laborers

BST90P25 Dispersion in occupation type for employed persons 16 years and over (managerial or professional/technical, sales or administrative support/service occupations/farming, forestry or fishing/production, craft or repair/operators, fabricators and laborers)

HOUSING

Housing Units

BST90P26 Tenure of occupied housing units

1=heavily renter occupied
2=mixed tenure
3=heavily owner occupied

Year Householder Moved into Unit

BST90P27 Proportion occupied housing units moved into between 1985 and March 1990[1]
1=low
2=medium
3=high

Housing Units

BST90P28 Median value of specified owner-occupied housing units (10 categories)[1]

BST90P29 Dispersion in value of specified owner-occupied housing units (10 categories)[1]

Notes

1. See APPENDIX B - Variable Category Determinations.

Appendix A - Statistical Measure Definitions

This technical appendix describes the statistical measures used to calculate contextual variables in the Add Health Public Use Contextual Database.

Dispersion

The dispersion measures are based on the following formula:

$$D = \frac{k(N^2 - \Sigma f_i^2)}{N^2(k-1)}$$

where k is the number of categories, N^2 is the sum of all categories squared, and Σf_i^2 is the sum of squared category frequencies over all i (=1,k) groups. If D = 0, only one category is nonzero; if D = 1, all category frequencies are equal.

Medians

Median values for grouped data were calculated using Pareto interpolation for income measures and linear interpolation for the age and housing value measures. The formula for the median is as follows:

$$M_{lb} + [p*(M_{ub} - M_{lb})]$$

where M_{ub} is the upper bound of the category containing the median, M_{lb} is the lower bound of this category, and p is the proportion of the population bounded by M_{ub} and M_{lb} that lies at or below the median. In Pareto interpolation, the median is derived by interpolating between the logarithms of the M_{ub} and M_{lb}.

If the median falls in the final open-ended interval of any distribution, the median is equated to the lower limit of this category minus one. For family and household income, this value is 100,001; for housing value, this value is 300,001. If the median falls in the lowest interval of any distribution, the median is equated to the upper bound of the category minus one. For family and household income, this value is 4,999; for housing value, this value is 14,999. This procedure is consistent with that used to calculate median measures in the U.S. Bureau of Census, Summary Tape File 3A.[1]

Notes

1. Bureau of the Census. 1992. *Census of Population and Housing, 1990; Summary Tape File 3 on CD-ROM Technical Documentation.* Washington, DC: Bureau of the Census.

Appendix B - Variable Category Determinations

This technical appendix details the category determinations made in constructing various contextual variables in the Add Health Public Use Contextual Database. Each was categorized based on the distributional and substantive characteristics of the measure. Information for relevant variables is provided below in the order that they appear in the Data Dictionary.

BST90P01 Urbanicity code

The urbanicity code distinguishes block groups that are in completely urbanized areas (BST90P01=1) from those that have any individuals living outside urbanized areas, in rural farm or rural nonfarm locations (BST90P01=2). This measure is different from the census "urban" designation which also includes places outside urbanized areas of 2,500 or more persons. The urbanicity code was used in determining whether respondent residence matched geocodes, based on ZIP+4 centroids, were adequate identifiers of the residence block group (see Missing Data in the Introduction).

BST90P04 Proportion Hispanic

Multiple categories of the proportion Hispanic measure provide concentration detail of the Hispanic population in a block group. The "low" category (BST90P04=1) consists of those block groups where less that 25 percent of the population was Hispanic; block groups with a 25-49 percent Hispanic population were coded "medium" (BST90P04=2); block groups with 50-74 percent Hispanic population were coded "high" (BST90P04=3); and block groups with a population that was 75 percent or more Hispanic were coded as "very high" (BST90P04=4).

BST90P05 Sex composition

Sex composition categories are based on the distribution of the proportion female in the population.[1] Heavily male, balanced, and heavily female categories were determined by taking one standard deviation below and above the mean of the distribution of this measure. Block groups less than 47 percent female were coded "heavily male" (BST90P05=1); block groups between 47 and 56 percent female were coded as "balanced" (BST90P05=2); and block groups greater than 56 percent female were coded as "heavily female" (BST90P05=3).

BST90P10 Proportion population that are children under five years old

The distinction between low, medium and high proportions of the population comprised of children under five years old was determined by taking one standard deviation below and above the mean of the distribution.[1] Block groups where less than 4.3 percent of the population was under five years old were coded as "low" (BST90P10=1); block groups where this proportion was between 4.3 and 11 percent were coded "medium" (BST90P10=2); and those block groups where this proportion was greater than 11 percent were coded "high" (BST90P10=3).

BST90P15 - BST90P18 Household income and family income measures

Medians and dispersion measures of household and family income were calculated using nine aggregate income categories: less than $5,000; $5,000 to $9,999; $10,000 to $14,999; $15,000 to $24,999; $25,000 to $34,999; $35,000 to $49,999; $50,000 to $74,999; $75,000 to $99,999; and $100,000 or more.

BST90P19 Proportion of persons with income in 1989 below poverty level

Low, medium and high categories of low poverty concentration are based on the distribution of proportion of persons below poverty level in 1989.[1] Block groups where the proportion of the population with income below poverty level was less than 11.6 percent, the median proportion, were coded "low" (BST90P19=1); block groups where this proportion was between 11.6 and 23.9 percent were coded "medium" (BST90P19=2); and those block groups where this proportion was greater than 23.9 percent, or block groups among the highest 25 percent in low poverty, were coded "high" (BST90P19=3).

BST90P22 Proportion females aged 16 years and over in the civilian labor force
Low, medium, and high female labor force participation distinctions were determined by taking one standard deviation below and above the mean of this distribution.[1] Block groups where less than 44.3 percent of the population of females aged 16 and over were in the civilian labor force were coded "low" (BST90P22=1); block groups where this proportion was between 44.3 and 68.5 percent were coded "medium" (BST90P22=2); and block groups where this proportion was greater than 68.5 percent were coded as "high" (BST90P22=3).

BST90P23 Unemployment rate
Block groups with an unemployment rate less than 6.5 percent, the median rate, were coded "low" (BST90P23=1); those with rates between 6.5 and 10.9 percent were coded "medium" (BST90P23=2); and block groups with unemployment rates greater than 10.9 percent, comprised of those block groups among the top 25 percent in unemployment, were coded "high" (BST90P23=3).[1]

BST90P26 Tenure of occupied housing units
Housing unit tenure categories provide detail concerning the proportion of occupied housing units that are owner occupied. The "heavily renter occupied" category (BST90P26=1) consists of those block groups where less than 25 percent of the housing units were owner occupied; block groups with a 25 to 75 percent owner occupied population of housing units were coded "mixed tenure" (BST90P26=2); and block groups where more than 75 percent of the housing units were owner occupied were coded "heavily owner occupied" (BST90P26=3).

BST90P27 Proportion occupied housing units moved into between 1985 and March 1990
For a measure of the proportion of occupied housing units moved into between 1985 and March 1990, low, medium, and high distinctions were determined by taking one standard deviation below and above the mean of this distribution.[1] Block groups where less than 30.4 percent of the occupied housing units were moved into between 1985 and March 1990 were coded "low" (BST90P27=1); block groups where this proportion was between 30.4 and 65.0 percent were coded "medium" (BST90P27=2); and block groups where this proportion was greater than 65.0 percent were coded as "high" (BST90P27=3).

BST90P28 - BST90P29 Value of specified owner-occupied housing unit measures
The median and dispersion in specified owner-occupied housing unit value were calculated using ten housing value categories: less than $15,000; $15,000 to $24,999; $25,000 to $49,999; $50,000 to $74,999; $75,000 to $99,999; $100,000 to $149,999; $150,000 to $199,999; $200,000 to $249,999; $250,000 to $299,999; and $300,000 or more.

Notes

1. Distributional characteristics are based on the sample of Add Health respondent residence block groups.

Appendix C - Contextual Database Codebook

This technical appendix provides a codebook for the Wave I and Wave II data files that comprise the Add Health Public Use Contextual Database. Summary statistics and missing data frequencies are listed for each variable in the order that the variables reside in each data file. The Wave I Respondent Residence Data Codebook begins on page 14. It is followed by the Wave II Respondent Residence Data Codebook that begins on page 27. Both codebooks and files have identical structures.

Frequency	Code	Response	Variable Name	Type/ Length
Respondent AID			**AID**	char 8
Geocode Match Indicator			**MATCH**	num 1
66	0	no match		
1407	1	GPS match		
3277	2	address match		
84	3	ZIP+4 match/urban		
Mover Indicator: Respondent Moved Between Wave I and Wave II			**MOVER**	num 1
4533	0	respondent did not move		
214	1	moved to different block group		
39	2	moved within same block group		
48	3	moved, location unknown		
Urbanicity Code			**BST90P01**	num 1
2443	1	completely urban		
2320	2	not completely urban		
5	8	unstable estimates		
66	9	geocode missing		
Modal Race			**BST90P02**	num 1
3880	1	white		
698	2	black		
182	3	other		
8	8	unstable estimates		
66	9	geocode missing		
Dispersion in Race Composition			**BST90P03**	num 4
4760		range 0 to 0.997		
8	9998	unstable estimates		
66	9999	geocode missing		
Proportion Hispanic			**BST90P04**	num 1
4349	1	low		

Frequency	Code	Response	Variable Name	Type/ Length
223	2	medium		
105	3	high		
86	4	very high		
5	8	unstable estimates		
66	9	geocode missing		
Sex Composition			**BST90P05**	num 1
450	1	heavily male		
3882	2	balanced		
431	3	heavily female		
5	8	unstable estimates		
66	9	geocode missing		
Median Age			**BST90P06**	num 2
4760		range 17 to 72		
8	98	unstable estimates		
66	99	geocode missing		
Dispersion in Age Distribution			**BST90P07**	num 4
4760		range 0.334 to 0.998		
8	9998	unstable estimates		
66	9999	geocode missing		
Modal Marital Status			**BST90P08**	num 4
465	1	never married		
4269	2	married, spouse present		
15	3	separated or divorced		
19	8	unstable estimates		
66	9	geocode missing		
Dispersion in Marital Status			**BST90P09**	num 4
4749		range 0 to 0.999		
19	9998	unstable estimates		

Frequency	Code	Response	Variable Name	Type/ Length
66	9999	geocode missing		
Proportion of Population that are Children Under Five Years Old			**BST90P10**	num 1
530	1	low		
3745	2	medium		
485	3	high		
8	8	unstable estimates		
66	9	geocode missing		
Modal Migration Status			**BST90P11**	num 1
4093	1	lived in same house in 1985		
292	2	lived in different house in 1985, same county		
375	3	lived in different house in 1985, different county		
8	8	unstable estimates		
66	9	geocode missing		
Dispersion in Migration Status			**BST90P12**	num 4
4760		range 0.131 to 1		
8	9998	unstable estimates		
66	9999	geocode missing		
Modal Household Type			**BST90P13**	num 1
3990	1	married couple family household		
209	2	other family household		
417	3	non-family household		
152	8	unstable estimates		
66	9	geocode missing		
Dispersion in Household Type			**BST90P14**	num 4
4616		range 0.121 to 1		
152	9998	unstable estimates		
66	9999	geocode missing		
Median Household Income in 1989			**BST90P15**	num 6

Frequency	Code	Response	Variable Name	Type/ Length
4616		range $4,999 to $100,001		
152	999998	unstable estimates		
66	999999	geocode missing		
Dispersion in Household Income in 1989			**BST90P16**	num 4
4616		range 0.494 to 0.987		
152	9998	unstable estimates		
66	9999	geocode missing		
Median Family Income in 1989			**BST90P17**	num 6
4309		range $4,999 to $100,001		
459	999998	unstable estimates		
66	999999	geocode missing		
Dispersion in Family Income in 1989			**BST90P18**	num 4
4309		range 0.409 to 0.986		
459	9998	unstable estimates		
66	9999	geocode missing		
Proportion Persons with Below Poverty-Level Income in 1989			**BST90P19**	num 1
2641	1	low		
1076	2	medium		
1046	3	high		
5	8	unstable estimates		
66	9	geocode missing		
Modal Educational Attainment of Individuals Aged 25 Years and Over			**BST90P20**	num 1
752	1	no high school degree or equivalency		
3514	2	high school degree, no college degree		
475	3	college degree or more		
27	8	unstable estimates		
66	9	geocode missing		

Frequency	Code	Response	Variable Name	Type/ Length
Dispersion in Educational Attainment of Individuals Aged 25 Years and Over			**BST90P21**	num 4
4741		range 0.159 to 1		
27	9998	unstable estimates		
66	9999	geocode missing		
Proportion Females Aged 16 Years and Over in Civilian Labor Force			**BST90P22**	num 1
805	1	low		
3126	2	medium		
729	3	high		
108	8	unstable estimates		
66	9	geocode missing		
Unemployment Rate			**BST90P23**	num 1
2506	1	low		
1138	2	medium		
1028	3	high		
96	8	unstable estimates		
66	9	geocode missing		
Modal Occupation Type for Employed Person 16 Years and Over			**BST90P24**	num 1
1041	1	managerial or professional		
2439	2	technical, sales or administrative support		
281	3	service occupations		
34	4	farming, forestry or fishing		
74	5	production, craft or repair		
763	6	operators, fabricators and laborers		
136	8	unstable estimates		
66	9	geocode missing		
Dispersion in Occupation Type for Employed Persons 16 Years and Over			**BST90P25**	num 4
4632		range 0.326 to 0.994		

Frequency	Code	Response	Variable Name	Type/ Length
136	9998	unstable estimates		
66	9999	geocode missing		
Tenure of Occupied Housing Units			**BST90P26**	num 1
272	1	heavily renter occupied		
2137	2	mixed tenure		
2346	3	heavily owner occupied		
13	8	unstable estimates		
66	9	geocode missing		
Proportion Occupied Housing Units Moved into Between 1985 and March 1990			**BST90P27**	num 1
709	1	low		
3327	2	medium		
573	3	high		
159	8	unstable estimates		
66	9	geocode missing		
Median Housing Value of Owner-Occupied Housing Units			**BST90P28**	num 6
3041		range $149,99 to $300,001		
1727	9998	unstable estimates		
66	9999	geocode missing		
Dispersion in Value of Specified Owner-Occupied Housing Units			**BST90P29**	num 4
3041		range 0 to 0.956		
1727	9998	unstable estimates		
66	9999	geocode missing		