

# General Best Practices for Using Add Health Data

---

ROBERT A. HUMMER

HOWARD W. ODUM DISTINGUISHED PROFESSOR OF SOCIOLOGY

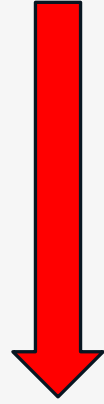
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

ADD HEALTH USERS CONFERENCE JUNE 17-18, 2024

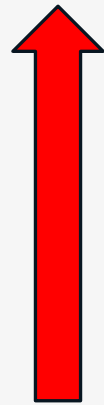
# Online Codebook Explorers

---

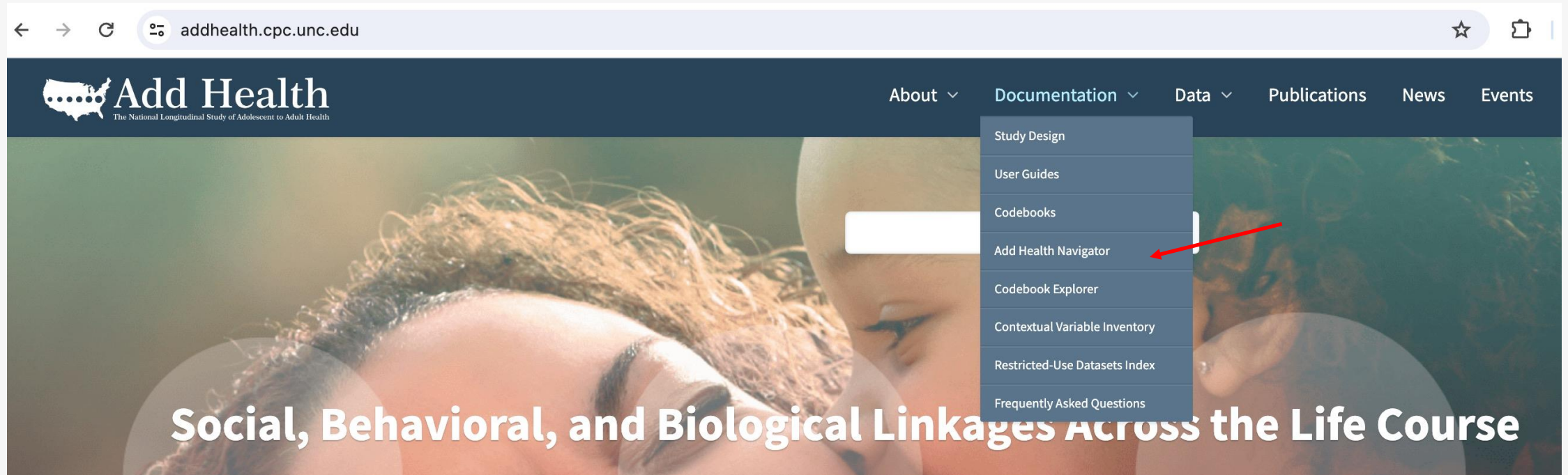
# Where to start?



<https://addhealth.cpc.unc.edu/>



# Add health navigator



The screenshot shows the website [addhealth.cpc.unc.edu](http://addhealth.cpc.unc.edu). The navigation bar includes links for About, Documentation, Data, Publications, News, and Events. The 'Documentation' dropdown menu is open, listing the following items: Study Design, User Guides, Codebooks, Add Health Navigator (highlighted with a red arrow), Codebook Explorer, Contextual Variable Inventory, Restricted-Use Datasets Index, and Frequently Asked Questions. The main banner features a photograph of a woman and a child with the text: "Social, Behavioral, and Biological Linkages Across the Life Course".

## Home

### Welcome

### Welcome to the Add Health Navigator

You are invited to explore the Add Health codebooks as well as the Wave I-V variable comparison without signing in. **If you would like to use the Baskets tab to create custom codebooks, you will need to sign in with a Microsoft account (personal or work or school).** To sign in, select "Log in" from the user account drop down menu located in the upper right hand corner of this page. If your work or school account will not allow you to grant permission to access resources, please start over and instead create a new Microsoft account. To create a new account, select "Create one!" from the Microsoft Sign In screen and follow the prompts. Return to this site once complete.

#### Some helpful tips:

- \* Click on the blue button below "Add Health Series" to view codebooks sorted by associated Wave.
- \* Click the "Search" button at the top of the page to search all datasets and questions for a particular keyword or phrase.
- \* Click "Explore" to explore Add Health Questions by topic and view a wave to wave comparison of the questions.



# Add Health

The National Longitudinal Study of Adolescent to Adult Health

## Explore

 Home / Explore

- [Add Health Topics](#)

Add Health Navigator Search Explore Baskets 0

Help

- Academics and Education
- Access to Health Services
  - BEM Inventory
- Biospecimen
- Census Region
- Children
- Civic Participation
- Cognitive
- Community Service
- Contraception
- Criminal Offending and Victimization
- Daily Activities
- Delinquency and Violence
- Demographics
- Development
- Diet and Nutrition
- Disposition
- Early Life
- Economics
- Employment
- Expectations
- Experiences
- Female Physical Development
- Field Interviewer

Select an item on the left.



- Interview/Questionnaire/Survey
- Male Physical Development
- Mental Health**
- Mentoring
- Military Service
- Mistreatment by Adults
- Motivations to Engage in Risky Behaviors
- Neighborhood
- Non-Relationship History
- Organizations
- Parental Relationships
- Parenting
- Personality, Attitudes, and Feelings
- Physical/Functional Limitations
- Pregnancy
- Relationships
- Religion and Spirituality
- Residence History
- School
- Self-Image
- Sexual Experiences
- Sexual Relationships
- Sexually Transmitted Diseases and Infections
- Siblings
- Social Network and Support

	Add Health						
	Wave I	Wave II	Wave III	Wave IV	Wave V	In-School	
	Wave I In-Home Interview	Parents (1995)	Wave II In-Home Interview	Wave III In-Home Interview	Wave IV In-Home Interview	Wave V Mixed-Mode Survey	In-School Questionnaire
	Wave I In-Home Interview Data	Parents (1995)	Wave II In-Home Interview	Wave III In-Home Interview	Wave IV In-Home Interview	Wave V Mixed-Mode Survey	Wave I In-School Questionnaire Data
How often have you cried frequently?	H1GH21		H2GH26	H3SP2			S60N
<a href="#">Have you ever been told by a doctor that you have an eating disorder, such as anorexia nervosa or bulimia?</a>				H3GH8		H5ID6T	
In the past 12 months, how often have you laughed a lot?				H3SP1			
In the last 30 days, how often have you felt that you were unable to control the important things in your life?					H4MH3	H5MN1	
In the last 30 days, how often have you felt confident in your ability to handle your personal problems?					H4MH4	H5MN2	



Add Health Navigator Search Explore Baskets 0

Wave I In-Home Interview Data (89 of 2240)

### Variable Description

Name H1GH21  
Label S3Q21 FREQ-FREQUENT CRYING-W1  
Dataset wave1

H1GH21

### Summary Statistics for 'H1GH21'

Value	Valid	Label	Frequency	% of Valid	% of All
0	✓	never	14,282	68.95%	68.85%
1	✓	just a few times	5,054	24.40%	24.36%
2	✓	about once a week	994	4.80%	4.79%
3	✓	almost every day	280	1.35%	1.35%
4	✓	every day	104	0.50%	0.50%
6	✗	refused	18		0.09%
8	✗	don't know	13		0.06%

Total Responses	Valid	Invalid	Min	Max
20745	20714	31	0	4

Appears Within

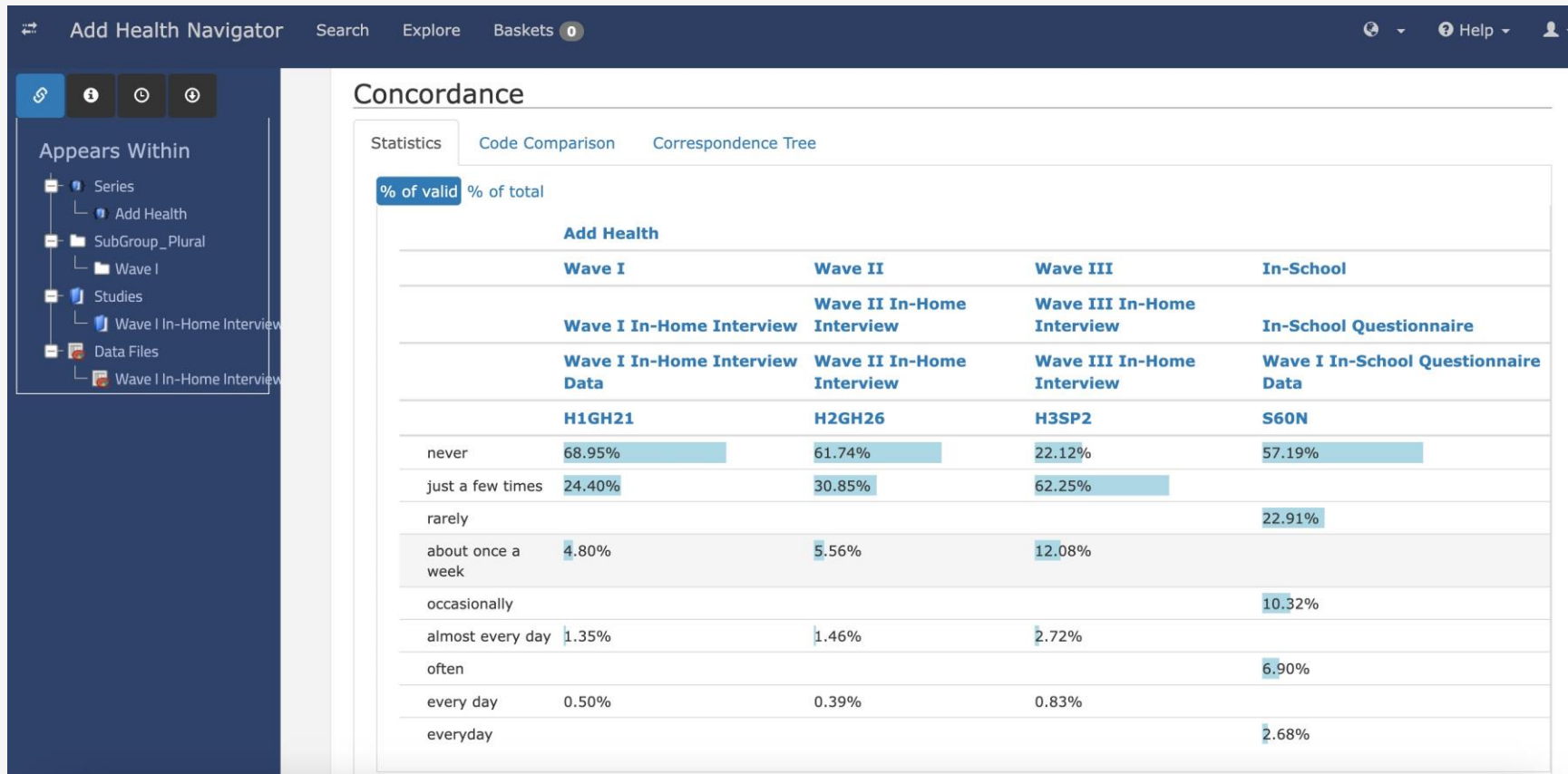
- Series
  - Add Health
- SubGroup\_Plural
  - Wave I
- Studies
  - Wave I In-Home Interview
- Data Files
  - Wave I In-Home Interview

## Representation

Representation Type	Code List
Selection Style	SelectOne
Measurement Unit	numeric
Codes	<input type="checkbox"/> never, just a few times, about once a week <ul style="list-style-type: none"><li>0 <input checked="" type="checkbox"/> never</li><li>1 <input checked="" type="checkbox"/> just a few times</li><li>2 <input checked="" type="checkbox"/> about once a week</li><li>3 <input checked="" type="checkbox"/> almost every day</li><li>4 <input checked="" type="checkbox"/> every day</li></ul>
Measurement Unit	numeric
Aggregation Method	Unspecified
Temporal	False
Geographic	False
Represented Variable	<input type="checkbox"/> How often have you cried frequently?
Missing Values	<input type="checkbox"/> refused, don't know

## Source Questions

H1GH21



## Home

### Welcome

### Welcome to the Add Health Navigator

You are invited to explore the Add Health codebooks as well as the Wave I-V variable comparison without signing in. **If you would like to use the Baskets tab to create custom codebooks, you will need to sign in with a Microsoft account (personal or work or school).** To sign in, select "Log in" from the user account drop down menu located in the upper right hand corner of this page. If your work or school account will not allow you to grant permission to access resources, please start over and instead create a new Microsoft account. To create a new account, select "Create one!" from the Microsoft Sign In screen and follow the prompts. Return to this site once complete.

#### Some helpful tips:

- \* Click on the blue button below "Add Health Series" to view codebooks sorted by associated Wave.
- \* Click the "Search" button at the top of the page to search all datasets and questions for a particular keyword or phrase.
- \* Click "Explore" to explore Add Health Questions by topic and view a wave to wave comparison of the questions.



# Add Health

The National Longitudinal Study of Adolescent to Adult Health

Add Health Series

Search

Search



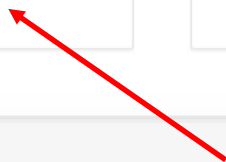
Search



231  
Data Files



31,741  
Variables



# Data files

The screenshot shows the 'Add Health Navigator' search interface. The top navigation bar includes 'Add Health Navigator', 'Search' (highlighted), 'Explore', and 'Baskets' with a count of 0. On the right side of the navigation bar are icons for a globe, 'Help', and a user profile. The main content area is titled 'Data Files'. It features a search input field with the placeholder text 'Search', a 'Sort by:' dropdown menu currently set to 'Default', and an 'Advanced...' dropdown menu. Below these is a blue 'Search' button. Underneath the search controls, it says 'Item types: Data Files'. The results section shows 'Results 1 to 50 of 231 (0.03 seconds)'. Two data file entries are visible, each with a document icon and a title: 'Baroreflex Sensitivity and Hemodynamic Recovery' and 'Measures of Inflammation and Immune Function'. Each entry is followed by a white rectangular area, likely representing a description or details for that file.

Add Health Navigator Search Explore Baskets 0

Wave I In-Home Interview Data

Wave I In-Home Interview Data > Add Health > Wave I > Wave I In-Home Interview

Search within

Appears Within

- Metadata Packages
  - Wave I In-Home Interview
- Series
  - Add Health
- SubGroup\_Plural
  - Wave I
- Studies
  - Wave I In-Home Interview

**Dataset Description**

Title Wave I In-Home Interview Data

Case Quantity 20,745

Variables 2,240

Explore [Explore Dataset](#)

**Data**

Variables in 'Wave I In-Home Interview Data'

Name	Label
IMONTH	MONTH OF INTERVIEW-W1
IDAY	DAY OF INTERVIEW-W1
IYEAR	YEAR OF INTERVIEW-W1
SCID	SCHOOL ID-W1
SSCID	SISTER SCHOOL ID-W1
COMMID	COMMUNITY ID-W1
MACNO	INTERVIEWER MACHINE NUMBER-W1

## Variable Description

Name H1GI14  
 Label S1Q14 BORN A US CITIZEN-W1  
 Dataset [wave1](#)

H1GI14

### Summary Statistics for 'H1GI14'

Value	Valid	Label	Frequency	% of Valid	% of All	
0	✓	no	1,655	87.43%	7.98%	
1	✓	yes	238	12.57%	1.15%	
6	✗	refused	1		0.00%	
7	✗	legitimate skip	18,842		90.83%	
8	✗	don't know	9		0.04%	
<b>Total Responses</b>			<b>Valid</b>	<b>Invalid</b>	<b>Min</b>	<b>Max</b>
20745			1893	18852	0	1



## Concordance

Statistics Code Comparison Correspondence Tree

% of valid % of total

[Add Health](#)

	Wave I	Wave III	Wave IV
	<b>Wave I In-Home Interview</b>	<b>Wave III In-Home Interview</b>	<b>Wave IV In-Home Interview</b>
	<b>Wave I In-Home Interview Data</b>	<b>Wave III In-Home Interview</b>	<b>Wave IV In-Home Interview</b>
	<b>H1GI14</b>	<b>H3OD16</b>	<b>H4OD4</b>
no	87.43%	87.35%	6.23%
yes	12.57%	12.65%	93.77%

Dataset	Variable	Valid	Invalid	Min	First Quartile	Median	Third Quartile	Max	Mean	StdDev
Wave I In-Home Interview Data	<b>H1GI14</b>	1,893	18,852	0				1		
Wave III In-Home Interview	<b>H3OD16</b>	1,225	13,972	0				1		
Wave IV In-Home Interview	<b>H4OD4</b>	15,700	1	0				1		

## Add Health Series

Search



Search



231  
Data Files



31,741  
Variables



# Variables

The screenshot shows the 'Add Health Navigator' search interface. The top navigation bar includes 'Add Health Navigator', 'Search', 'Explore', and 'Baskets 0'. The main content area is titled 'Variables' and features a search input field containing 'born citizen', a 'Sort by: Default' dropdown, and an 'Advanced...' dropdown. A blue 'Search' button is positioned below the input field. The search results are displayed under the heading 'Item types: Variables' and 'Query: born citizen'. The results section shows 'Results 1 to 5 of 5 (0.12 seconds)'. Two results are visible: 'H1GI14' with the variable name 'S1Q14 BORN A US CITIZEN-W1', and 'H3OD16' with the variable name 'S1Q16 BORN A US CITIZEN-W3'.

Add Health Navigator Search Explore Baskets 0

Variables

born citizen Sort by: Default Advanced...

Search

Item types:  
Variables

Query:  
born citizen

Results 1 to 5 of 5 (0.12 seconds)

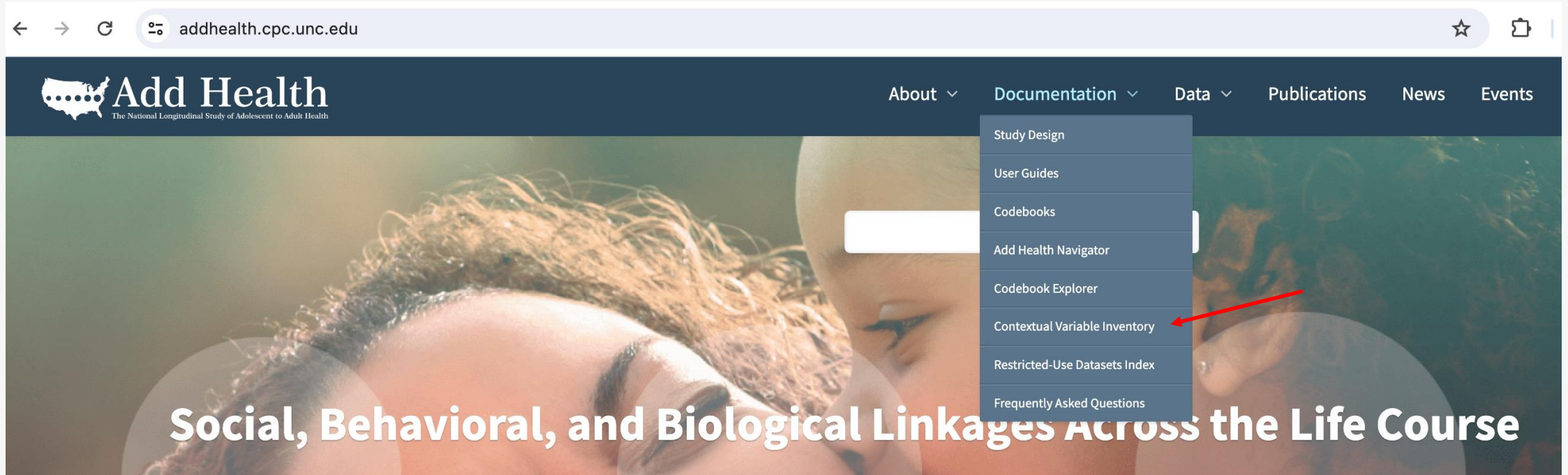
H1GI14  
S1Q14 BORN A US CITIZEN-W1

H3OD16  
S1Q16 BORN A US CITIZEN-W3

Let's walk through an example...

# Example

# Contextual variable inventory



The screenshot shows the Add Health website interface. The browser address bar displays 'addhealth.cpc.unc.edu'. The main navigation bar includes links for 'About', 'Documentation', 'Data', 'Publications', 'News', and 'Events'. A dropdown menu is open under 'Documentation', listing several options: 'Study Design', 'User Guides', 'Codebooks', 'Add Health Navigator', 'Codebook Explorer', 'Contextual Variable Inventory', 'Restricted-Use Datasets Index', and 'Frequently Asked Questions'. A red arrow points to the 'Contextual Variable Inventory' option. The background features a large image of a woman's face with the text 'Social, Behavioral, and Biological Linkages Across the Life Course' overlaid.

## Add Health Contextual Variable Inventory

The Add Health Contextual Variable Inventory is designed to allow researchers to browse or search through the available variables included in all Add Health contextual datasets.

Typing in the text boxes above the Variable Label and Variable Name column will filter as you type. Dropdown menus are available for most other columns. Lastly, we provide checkboxes that will display variables related to five topical domains of particular interest to Add Health.

Search:  Clear Search Reset To Default Sorting Show 10 entries

Variable Label	Wave	Category	Variable Name	Contextual Dataset	Data Source
----------------	------	----------	---------------	--------------------	-------------

The Add Health Contextual Variable Inventory is designed to allow researchers to browse or search through the available variables included in all Add Health contextual datasets.

Typing in the text boxes above the Variable Label and Variable Name column will filter as you type. Dropdown menus are available for most other columns. Lastly, we provide checkboxes that will display variables related to five topical domains of particular interest to Add Health.

 Search:  ✕



 Show  ▾ entries

 ▾

 ▾

 ▾

Variable Label	Wave	Category	Variable Name	Contextual Dataset	Data Source
HOLC REDLINING GRADE-W1	1	Politics and policies	W1HOLCGR	HOLC1345	Mapping Inequality: Redlining in New Deal A
HOLC REDLINING GRADE-W3	3	Politics and policies	W3HOLCGR	HOLC1345	Mapping Inequality: Redlining in New Deal A
HOLC REDLINING GRADE-W4	4	Politics and policies	W4HOLCGR	HOLC1345	Mapping Inequality: Redlining in New Deal A
PROXIMITY TO NEAREST HOLC AREA-W1	1	Politics and policies	W1HOLCPX	HOLC1345	Mapping Inequality: Redlining in New Deal A

Search:

[Clear Search](#)

[Reset To Default Sorting](#)

Show  entries

Select Geographic Scale

Select Raw/Constructed



**Geographic Scale**



**Raw/Constructed**



**Method/Algorithm**



**Structural Racism**



**Structural Sexism**



**Structural Xenophobia**



**Health**



**Heterosexism**



Participant proximity to redlined areas

Raw

See user guide

1

0

0

0

0

Participant proximity to redlined areas

Raw

See user guide

1

0

0

0

0

Participant proximity to redlined areas

Raw

See user guide

1

0

0

0

0



Let's walk through an example...

# Example

# Major Takeaways:

Once you've used the Add Health Navigator and/or Contextual Variable Inventory to find measures you may be interested in (Carlyn's presentation) and acquired the data:

- 1) Add Health has substantial online resources (e.g., User Guides) to help with analyses**
- 2) The complex Add Health survey design necessitates use of appropriate sampling weights to make population estimates**
- 3) Add Health has expertise to help with questions that may not be addressed online**

# Add Health Design and Analysis: Key Resources

1) Harris, et al. **Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health)**. *International Journal of Epidemiology* (published online June 29, 2019)

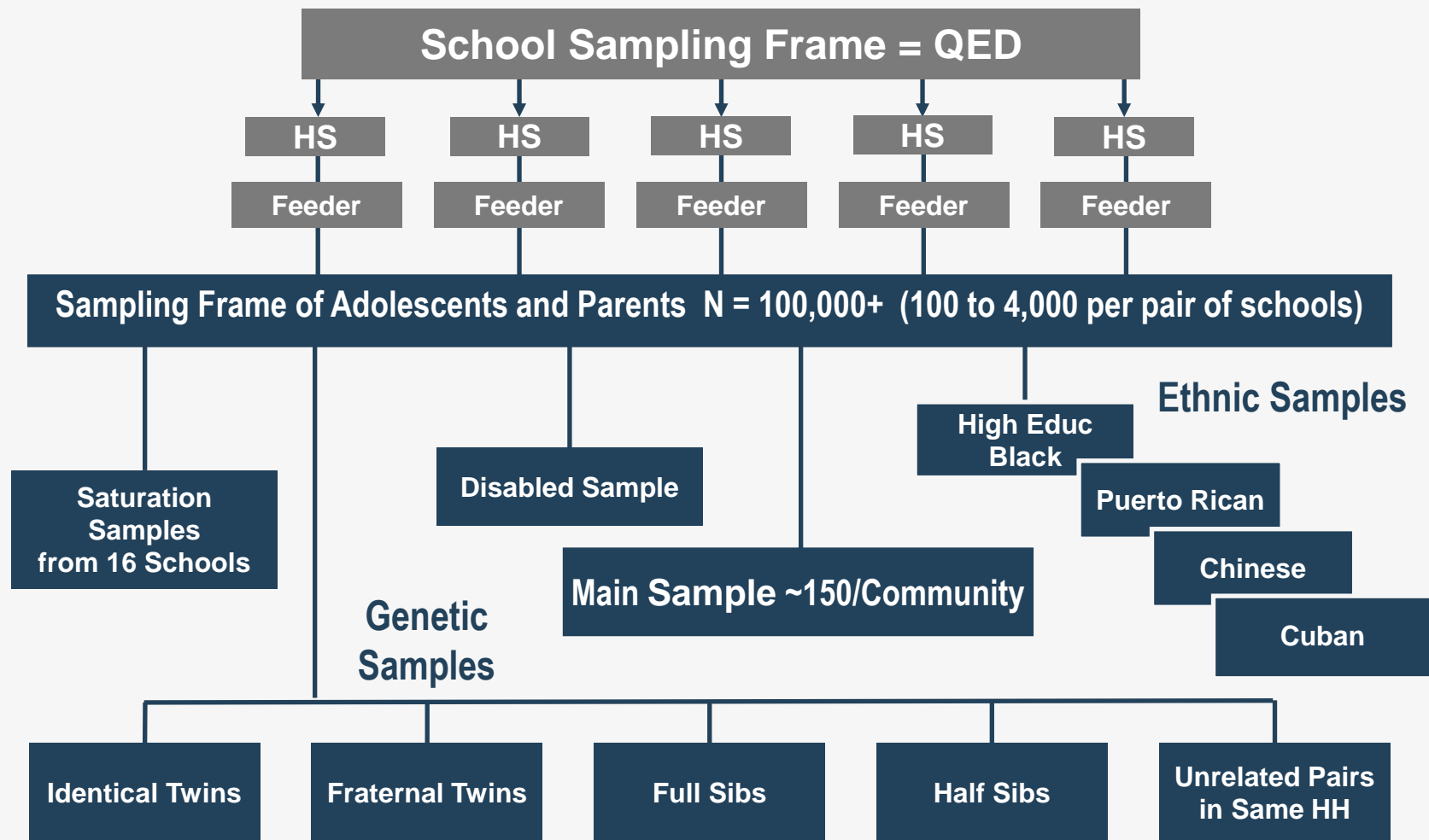
<https://doi.org/10.1093/ije/dyz115>

2) User Guides: <https://www.cpc.unc.edu/projects/addhealth/documentation/guides>

Chen, Ping, and Kathleen Mullan Harris. 2020. Guidelines for Analyzing Add Health Data. Carolina Population Center at the University of North Carolina at Chapel Hill.

[https://addhealth.cpc.unc.edu/wp-content/uploads/docs/user\\_guides/GuidelinesforAnalysisofAddHealthData\\_020422.pdf](https://addhealth.cpc.unc.edu/wp-content/uploads/docs/user_guides/GuidelinesforAnalysisofAddHealthData_020422.pdf)

Chen, Ping, and Kathleen Mullan Harris. 2020. Construction of Wave V Biomarker Weight. Carolina Population Center. [https://addhealth.cpc.unc.edu/wp-content/uploads/docs/user\\_guides/WaveVBiomarkerSampleWeightUserGuide.pdf](https://addhealth.cpc.unc.edu/wp-content/uploads/docs/user_guides/WaveVBiomarkerSampleWeightUserGuide.pdf)



# Understanding the Add Health Sample Design

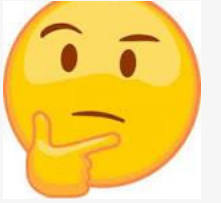
# Key Facets of Add Health Sampling Design

- **Clustering of individuals within schools**
  - Adolescents within the same school are not independent of one another
  - Adolescent outcomes will be more similar within schools than across schools
- **Unequal probability of sample selection**
- **Stratification by region**

# A Few Quirks of the Add Health Design

- 1) Not all Wave I participants have a sampling weight: 18,924 Wave I participants **have** a weight; 1,821 Wave I participants **do not have** a weight
- 2) A small number of new participants (**n = 27**) **were added in Wave II**; they do not have a sampling weight
- 3) Some Wave I participants who do not have a sampling weight **have not been followed up** since Wave II (n = 687)

# A Few Quirks, continued



4) Add Health has individual-based weights (part of design) and school-based weights (part of design), **but not neighborhood-based weights** (not part of design)

5) Wave V has 12,300 participants, most web-based; **among them is a nationally representative subset of in-person participants (N = 1,102). This subsample has its own weight variable to use: GSW5\_2B**

6) Wave V has a biomarker-based sample of 5,381, among whom 5,377 have weights. **This subsample has its own weight variable to use: W5BIOWGT**

# Sampling Design Adjustments

Design Attribute	Usual impact	Adjustment variable
Stratification	Reduce variance	Poststratification variable: census region
Clustering of students	Increase variance	PSU variable: School Identification
Unequal probability of selection	Increase variance; Biased parameter estimate	<p>Sampling Weights:</p> <ul style="list-style-type: none"> <li>• Cross-sectional weights for schools</li> <li>• Cross-sectional weights for analyzing each wave of data</li> <li>• Cross-sectional weights for analyzing sub-samples from Wave III and from Wave V</li> <li>• Longitudinal weights for conducting analysis combining data from multiple waves</li> <li>• Multilevel weights for two-level analysis where schools and individuals are levels of interest</li> </ul>



# Selection of Sampling Weights for Analysis

---

# Cross-sectional Analysis

Research questions that investigate an outcome measure(s) measured at one point in time

Predictor variables may be measured:

- At one wave
- At multiple waves

# Using Cross-Sectional Weights

X Variables	Y Variable	Weight
Same wave as Y	One Wave	Cross-sectional weight of Y variable (population-average model)
Multiple Waves	One Wave	Cross-sectional weight of Y variable (population-average model)

# Cross-Sectional Grand Sample Weights Single-Level (Population Average) Models

Data Set (Year collected)	Sampling Weight Variable (N)	Sample	Target Population
Wave I (1995)	GSWG1 (N=18,924)	Adolescents chosen with a known probability of being selected from 1994-1995 enrollment rosters of US schools	Adolescents in 1995 enrolled in grades 7-12 during the 1994-1995 academic year
Wave II (1996)	GSWG2 (N=13,570)	Wave I respondents who were interviewed at Wave II	Adolescents in 1996 enrolled in grades 7-12 during the 1994-1995 academic year
Wave III (2001)	GSWG3_2 (N=14,322)	Wave I respondents who were interviewed at Wave III	Adults in 2001 enrolled in grades 7-12 during the 1994-1995 academic year
Wave IV (2008)	GSWG4_2 (N=14,800)	Wave I respondents who were interviewed at Wave IV	Adults in 2008 enrolled in grades 7-12 during the 1994-1995 academic year
Wave V (2016-2018)	GSWG5 (N=12,300)	Wave I respondents who were interviewed at Wave V	Adults in 2016-18 enrolled in grades 7-12 during the 1994-1995 academic year

# Longitudinal Analysis

Research questions that investigate changes in an outcome measured for the same respondents over time

Outcome variable is measured multiple times

Predictor variables may be measured:

- At one wave
- At multiple waves

# Using Longitudinal Weights

X Variable	Y Variable	Weight
Wave I or multiple waves	Multiple waves for a sample of respondents who have data at every wave	Appropriate Longitudinal Weight for the multiple waves of your outcome (Y) variable

# Longitudinal Weights: Single-Level (Population Average) Models

Data Set (Year collected)	Sampling Weight Variable (N)	Sample	Target Population
Wave III (2001)	GSWG3 (N=10,828)	Eligible Wave I Respondents who were interviewed at both Wave II & Wave III	Adolescents enrolled in grades 7-12 during the 1994-1995 academic year interviewed in 1995, 1996 & 2001
Wave IV (2008)	GSWG4 (N=9,421)	Eligible Wave I respondents who were interviewed at Wave II, III & IV	Adolescents enrolled in grades 7-12 during 1994-1995 interviewed in 1995, 1996, 2001 & 2008
Wave IV (2008)	GSWG134 (N=12,288)	Eligible Wave I respondents who were interviewed at Wave III & IV	Adolescents enrolled in grades 7-12 during 1994-1995 interviewed in 1995, 2001 & 2008
Wave V (2016-2018)	GSWG12345 (N=7,295)	Eligible Wave I respondents who were interviewed at II, III, Wave IV & V	Adolescents enrolled in grades 7-12 during 1994-1995 interviewed in 1995, 1996, 2001, 2008 & 2016-18
Wave V (2016-2018)	GSWG1345 (N=9,349)	Eligible Wave I respondents who were interviewed at Wave III, IV & V	Adolescents enrolled in grades 7-12 during 1994-1995 interviewed in 1995, 2001, 2008 & 2016-18
Wave V (2016-2018)	GSWG145 (N=10,914)	Eligible Wave I respondents who were interviewed at Wave IV & V	Adolescents enrolled in grades 7-12 during 1994-1995 interviewed in 1995, 2008 & 2016-18

# Preparing Data for Analysis

- Determine the wave(s) of data
- Exclude cases with missing sampling weights
- Design type: Specify With Replacement as the Design Type
- Stratum Variable: **REGION**
- Cluster Variable or Primary Sampling Unit (PSU): **PSUSCID**



# Cross-sectional Weights Example

---



# Social Science & Medicine

Volume 314, December 2022, 115423



## Race/ethnicity, immigrant generation, and physiological dysregulation among U.S. adults entering midlife

Fatima Touma<sup>a b</sup>  , Robert A. Hummer<sup>a b</sup>

# Research Questions & Data

Wave V Biomarker data

- Q1 - How does the biological health of first- and second-generation Asian, Black, Hispanic, and White Americans compare to that of U.S.-born (3+ generation) individuals?


Wave I data

- Q2 - Are immigrant status and racial/ethnic differences in biological health explained by controls for childhood and adolescent social and health advantages and disadvantages?

Wave V data

- Q3 - Are immigrant status and racial/ethnic differences in biological health explained by controls for adulthood socioeconomic, social, and behavioral factors?

[Home](#) » [Documentation](#) » [User Guides](#)

## User Guides

Add Health produced the following detailed guides to assist analysts with understanding the data and performing unbiased analyses.

Open All Panels

Open All Panels and press Ctrl-F to search

### — General

#### [The Add Health Study: Design and Accomplishments](#)

by Kathleen Mullan Harris

Updated 2013

#### [Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health \(Add Health\)](#)

by Kathleen Mullan Harris, Carolyn Tucker Halpern, Eric A. Whitsel, Jon M. Hussey, Ley A. Killeya-Jones, Joyce Tabor, and Sarah C. Dean

2019

#### [Guidelines for Analyzing Add Health Data](#)

by Ping Chen and Kathleen Mullan Harris

Updated 2020



# Which Survey Weights to Use?

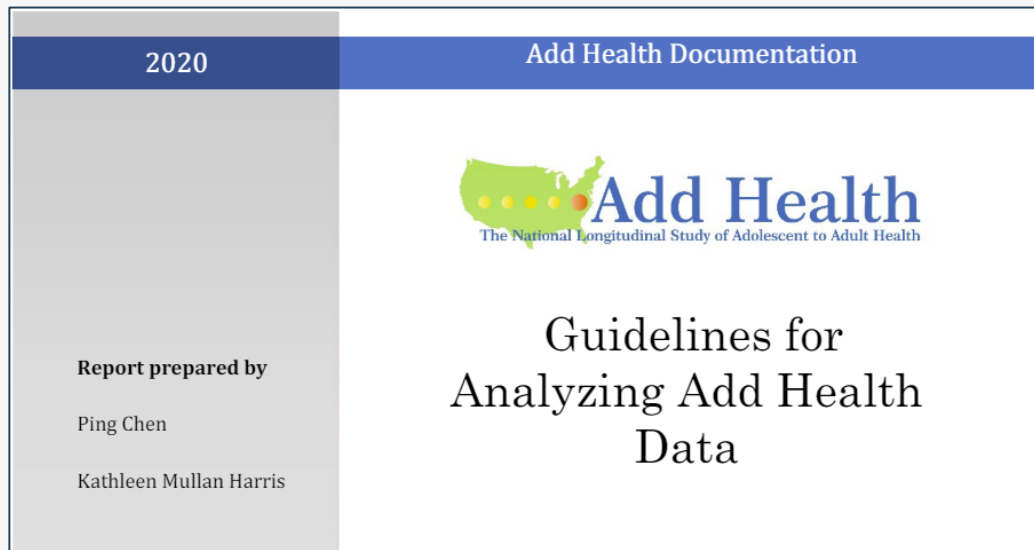
**Outcome:** Wave V biomarker measured  
allostatic load

## Covariates:

Wave I - self-rated health at Wave I, household structure, parent's education, and income

Wave V - respondent's age, race-ethnicity, sex, education, income, household structure, religious service attendance, smoking, and alcohol consumption.

**Population:** Individuals who attended middle or high school in 1994-1995 school year



**Table 2.4. Sampling Weights used in Cross-sectional Analysis**

Population of Interest	Data Used	Number of Participants in Analysis File	Sampling Weight Population Average Models	Sampling Weight Multilevel Models
Adolescents in 1995 enrolled in Grade 7-12 during 1994-1995	Wave I	18,924	GSWGT1	SCHWT1 W1_WC
Adolescents in 1996 enrolled in Grade 7-11 during 1994-1995	Wave II	13,570	GSWGT2	SCHWT1 W2_WC
Young Adults in 2001 enrolled in Grade 7-12 during 1994-1995	Wave III	14,322	GSWGT3_2	SCHWT1 W3_2_WC
Young adults Romantic Couplers in 2001 (one partner enrolled in Grade 7-12 during 1994-1995)	Wave III	1,317	W3PTNR	---
Young Adults in 2001 enrolled in Grade 7-12 during 1994-1995 (Educational analyses involving high school transcripts)	Wave III	11,637	TWGT3_2	---
Young Adults in 2001 enrolled in Grade 7-12 during 1994-1995 (Analyses involving special sample selected for testing urine for mycoplasma genitalium at Wave III.)	Wave III	14,322	MGENCRWT	---
Sexually Active Female Population	Wave III	6,593	HPVCRWT	---
Young Adults in 2008 enrolled in Grade 7-12 during 1994-1995	Wave IV	14,800	GSWGT4_2	SCHWT1 W4_2_WC
Young Adults in 2018 enrolled in Grade 7-12 during 1994-1995	Wave V	12,300	GSW5	---
Young Adults in 2018 enrolled in Grade 7-12 during 1994-1995	Wave V 2B Sample	1,102	GSW5_2B	---

**Table 2.8. Sampling Weights for Wave V Biomarker Sub-Sample with Single-Level Models**

Data Set (Year collected)	Sampling Weight Variable (N)	Sample	Target Population
Wave V (2018)	W5BLOWGT (N=5,377)	Biomarker sample with biomarker data assayed.	Young Adults in 2018 enrolled in Grade 7-12 during 1994-1995

## Codebooks

PDF codebooks for the Add Health restricted-use data

Open All Panels

Open All Panels and press Ctrl-F to search

- + Core Files
- + Friendship Files
- + Sibling Files
- + Contextual Data Files
- + Supplemental Files
- + Wave III Education Files
- Additional Weight Files



### Wave V Mixed-Mode Weights

The Wave V Mixed-Mode Weights were released with the Wave V Mixed-Mode Survey.

## Codebooks

PDF codebooks for the Add Health restricted-use data

Open All Panels

Open All Panels and press Ctrl-F to search

- + Core Files
- + Friendship Files
- + Sibling Files
- + Contextual Data Files
- + Supplemental Files
- + Wave III Education Files
- + Additional Weight Files
- + Network Files
- + Disposition Files
- + Wave I & III Obesity And Neighborhood Environment (ONE) Files
- + Wave III Biomarker Files
- + Wave IV Biomarker Files
- Wave V Biomarker Files



### Wave V Biomarker Weight

This file contains the Wave V biomarker sample weight.

# Finding Survey Weights

File: bweight5.xpt

AID – Respondent identifier

W5BIOWGT – Wave V biomarker sample weight

File: weights5.xpt

AID – Respondent identifier

PSUSCID – cluster variable-school ID

Region – Strata variable – region



# Merging Survey Weights to Data

```
/****** Merge Weights *****/

*Import wave V data:
import sasxport5 "D:\ahd\adhealth\wave5\wave5.xpt"
//Make a copy that will be your working dataset
save "mywave5.dta", replace

//You will also want to import and save a copy of the weights files
import sasxport5 "D:\ahd\adhealth\wave5\weights5.xpt"
save "W5_weights.dta", replace

import sasxport5 "D:\ahd\adhealth\wave5\bweight5.xpt"
save "W5_bioweight.dta", replace

*Merge weights to your dataset copy:
use "D:\...\mywave5.dta", clear
merge 1:1 aid using "D:\...\W5_weights.dta", keepusing(aid psuscid region
gsw5)

merge 1:1 aid using "D:\...\W5_bioweight.dta"
```

# Survey Design

```
/****** Account for Survey Design *****/  
svyset psuscid [pweight=w5biowgt], strata(region)
```

Output: `. svyset psuscid [pweight=w5biowgt], strata(region)`

```
Sampling weights: w5biowgt  
                  VCE: linearized  
                  Single unit: missing  
                  Strata 1: region  
                  Sampling unit 1: psuscid  
                  FPC 1: <zero>
```

# Subpopulation

```

/***** Subpopulation *****/
//Define subpopulation for complete case
analysis
gen subpop=0

replace subpop=1 if w5biowgt!=. & region!=. &
psuscid="" & racegen!=. & h1_health!=. &
h1_predu!=. & h1_hhincome!=. & h1_fam!=. &
h5_age!=. & h5_edu!=. & h5_hhincome!=. & h5_fam
!=. & h5_attend!=. & h5_smoke!=. & h5_alcoh!=. &
preg!=1 & allostatic_load!=. //n=2,946

```

ID	V1	V2	V3	V4	V5	V6	nmis
1	0	1	2	1	2	0	1
2	1	2	1	0	3	1	1
3	1	3	3	0	1	1	1
4	0	3	4	1	1	0	1
5	.	2	3	.	2	.	0
6	1	.	4	1	.	1	0
7	0	1	.	.	2	0	0
8	0	3	2	0	2	0	1
9	1	1	1	1	3	1	1
10	1	2	4	0	1	0	1

# Analysis

```
/****** Table 1 - Complete-case Descriptives *****/  
  
*Mean AL by race-ethnicity and immigrant gen  
svy, subpop(subpop): mean allostatic_load, over(racegen)  
  
*Mean age by race-ethnicity and immigrant gen  
svy, subpop(subpop): mean h5_age, over(racegen)  
  
*Proportion female by race-ethnicity and immigrant gen  
svy, subpop(subpop): proportion female, over(racegen)
```

```
. svy, subpop(subpop): mean allostatic_load, over(racegen)
(running mean on estimation sample)
```

Survey: Mean estimation

Number of strata = 4  
 Number of PSUs = 132

Number of obs = 5,269  
 Population size = 20,179,594  
 Subpop. no. obs = 2,946  
 Subpop. size = 11,218,619  
 Design df = 128

	Linearized			
	Mean	std. err.	[95% conf. interval]	
c.allostatic_load@racegen				
Asian Gen 1	2.280417	.5960844	1.100962	3.459872
Asian Gen 2	4.085469	.660981	2.777606	5.393333
Asian Gen 3+	3.323457	.4977202	2.338633	4.308282
Hispanic Gen 1	3.661451	.5804894	2.512853	4.810048
Hispanic Gen 2	3.505871	.3178848	2.876882	4.13486
Hispanic Gen 3+	4.394187	.827862	2.75612	6.032253
Black Gen 1	5.184975	.4905078	4.214421	6.155528
Black Gen 2	4.60192	.694615	3.227506	5.976335
Black Gen 3+	4.709844	.187492	4.338859	5.080829
White Gen 1	2.113395	.614168	.8981586	3.328631
White Gen 2	4.176376	.3084081	3.566138	4.786614
White Gen 3+	4.060061	.1020467	3.858144	4.261978

```
. mean allostatic_load if subpop==1, over(racegen)
```

Mean estimation

Number of obs = 2,946

	Mean	Std. err.	[95% conf. interval]	
c.allostatic_load@racegen				
Asian Gen 1	2.75	.4869324	1.795238	3.704762
Asian Gen 2	3.584906	.3861104	2.827832	4.341979
Asian Gen 3+	3.25	.5374158	2.196251	4.303749
Hispanic Gen 1	3.578947	.3548336	2.8832	4.274694
Hispanic Gen 2	3.903704	.2420372	3.429125	4.378283
Hispanic Gen 3+	4.097561	.3286069	3.453239	4.741883
Black Gen 1	4.25	1.108678	2.076138	6.423862
Black Gen 2	4.181818	.5715054	3.061228	5.302409
Black Gen 3+	4.436242	.1394694	4.162774	4.709709
White Gen 1	3	.700649	1.626188	4.373812
White Gen 2	3.711712	.2436373	3.233995	4.189428
White Gen 3+	3.830864	.0669875	3.699516	3.962211

```
. svy, subpop(subpop): mean allostatic_load, over(racegen)
(running mean on estimation sample)
```

Survey: Mean estimation

Number of strata = 4

Number of PSUs = 132

Number of obs = 5,269

Population size = 20,179,594

Subpop. no. obs = 2,946

Subpop. size = 11,218,619

Design df = 128

	Linearized Mean	std. err.	[95% conf. interval]	
c.allostatic_load@racegen				
Asian Gen 1	2.280417	.5960844	1.100962	3.459872
Asian Gen 2	4.085469	.660981	2.777606	5.393333
Asian Gen 3+	3.323457	.4977202	2.338633	4.308282
Hispanic Gen 1	3.661451	.5804894	2.512853	4.810048
Hispanic Gen 2	3.505871	.3178848	2.876882	4.13486
Hispanic Gen 3+	4.394187	.827862	2.75612	6.032253
Black Gen 1	5.184975	.4905078	4.214421	6.155528
Black Gen 2	4.60192	.694615	3.227506	5.976335
Black Gen 3+	4.709844	.187492	4.338859	5.080829
White Gen 1	2.113395	.614168	.8981586	3.328631
White Gen 2	4.176376	.3084081	3.566138	4.786614
White Gen 3+	4.060061	.1020467	3.858144	4.261978

# Subpopulation

Table 2.8. Sampling Weights for Wave V Biomarker Sub-Sample with Single-Level Models

Data Set (Year collected)	Sampling Weight Variable (N)	Sample	Target Population
Wave V (2018)	W5BIOWGT (N=5,377)	Biomarker sample with biomarker data assayed.	Young Adults in 2018 enrolled in Grade 7-12 during 1994-1995

```
. count if w5biowgt!=. & region!=. & psuscid=""  
5,269
```

# Analysis

```

/***** Model 1 *****/
*Model 1: Race-ethnicity immigrant generation, Sex, and Age
svy, subpop(subpop): nbreg allostatic_load ib12.racegen i.female
c.h5_age, irr
  
```

```

//You can also subset using an if statement
svy, subpop(subpop if female==1): nbreg allostatic_load ib3.racegen
i.female c.h5_age, irr
  
```

```

. *Model 1: Race-ethnicity immigrant generation, Sex, and Age
. svy, subpop(subpop): nbreg allostatic_load ib12.racegen i.female c.h5_age, irr
(running nbreg on estimation sample)
  
```

Survey: Negative binomial regression

Number of strata = 4  
Number of PSUs = 132

Number of obs = 5,269  
Population size = 20,179,594  
Subpop. no. obs = 2,946  
Subpop. size = 11,218,619  
Design df = 128  
F(13, 116) = 7.36  
Prob > F = 0.0000

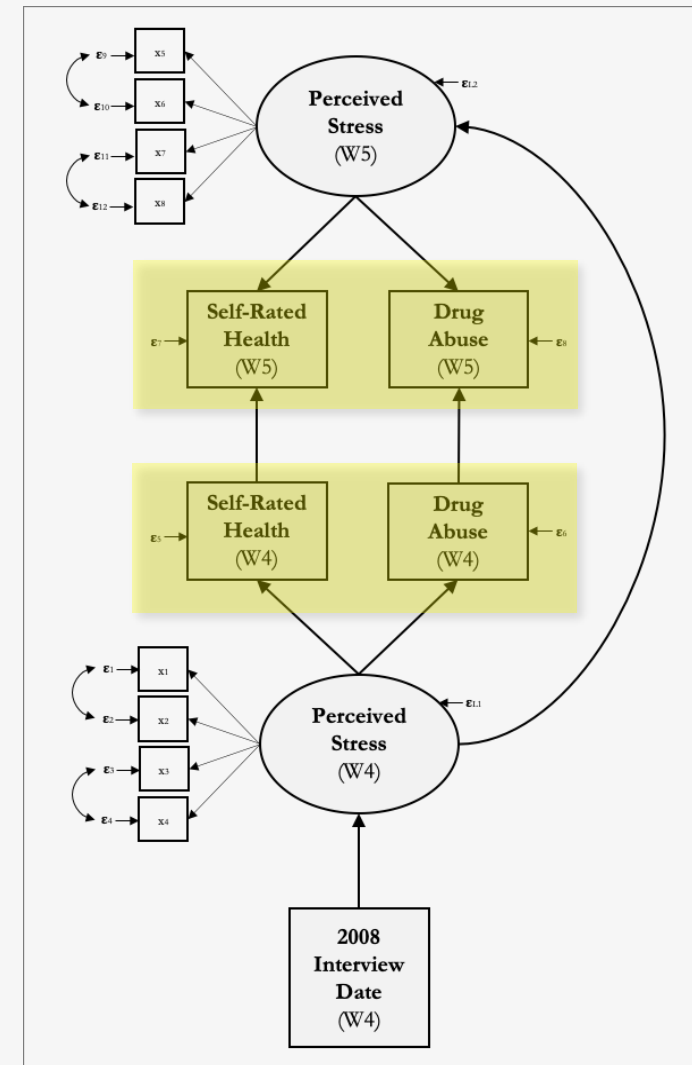
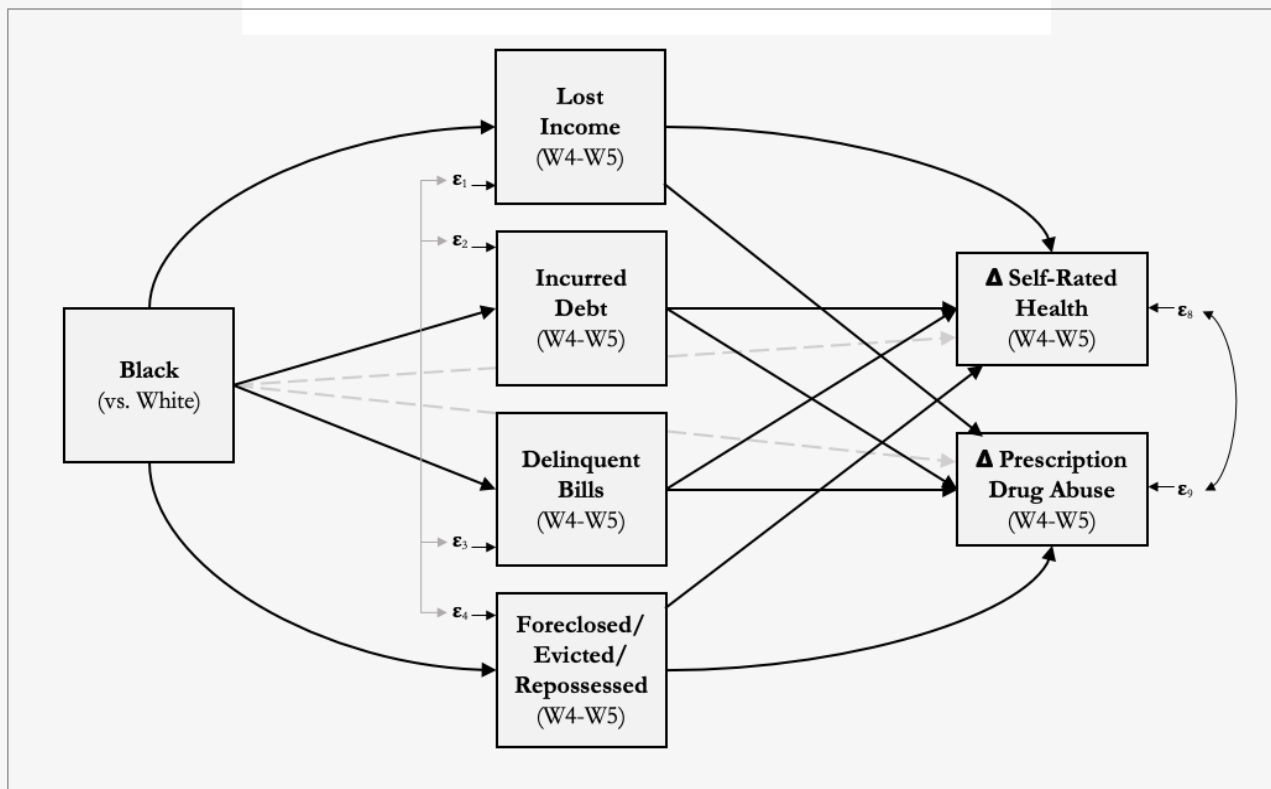
Dispersion: mean

allostatic_load	Linearized		t	P> t	[95% conf. interval]	
	IRR	std. err.				
racegen						
Asian Gen 1	.5336595	.1519956	-2.20	0.029	.3037478	.9375951
Asian Gen 2	1.046037	.195711	0.24	0.810	.7223861	1.514694
Asian Gen 3+	.8857243	.0997259	-1.08	0.283	.7088346	1.106757
Hispanic Gen 1	.8523746	.1286054	-1.06	0.292	.6323774	1.148907
Hispanic Gen 2	.861571	.0805091	-1.59	0.113	.7161299	1.03655
Hispanic Gen 3+	1.061726	.1904176	0.33	0.739	.7445513	1.514015
Black Gen 1	1.095884	.1214926	0.83	0.410	.8800299	1.364683
Black Gen 2	1.159288	.1696549	1.01	0.314	.8678291	1.548633
Black Gen 3+	1.164532	.0557416	3.18	0.002	1.0593	1.280218
White Gen 1	.5447151	.1689544	-1.96	0.052	.2948711	1.006252
White Gen 2	1.048345	.0766601	0.65	0.520	.9071225	1.211552
1.female						
h5_age	.7273125	.0358285	-6.46	0.000	.6597652	.8017754
_cons	1.027952	.0093286	3.04	0.003	1.009659	1.046577
1.female	1.681232	.5687165	1.54	0.127	.8608753	3.283333
/lnalpha	-1.240745	.0678524			-1.375002	-1.106487
alpha	.2891688	.0196208			.252839	.3307186



# Longitudinal Weights Example

---



```
/****** STEP 1: CREATE DATASET (AHTS) *****/

/// Wave I Survey ///
fdause "D:\ahd\adhealth\wave1\allwave1.xpt", clear
keep aid // etc.
save "wave1.dta", replace

/// Wave IV Survey ///
fdause "D:\ahd\adhealth\wave4\wave4.xpt", clear
keep aid // etc.
save "wave4.dta", replace

/// Wave V Survey ///
fdause "D:\ahd\adhealth\wave5\wave5.xpt", clear
keep aid // etc.
save "wave5.dta", replace

/// Longitudinal Weights ///
fdause "D:\ahd\adhealth\wave5\weights5.xpt", clear
keep aid psuscid region gsw145
save "weights.dta", replace
```

```
/****** STEP 1: CREATE DATASET (AHTS) *****/  
  
/// Merge Files ///  
use "wave1.dta", clear  
merge 1:1 aid using "wave4.dta", nogen  
merge 1:1 aid using "wave5.dta", nogen  
merge 1:1 aid using "weights.dta", nogen  
save "wave145.dta", replace  
  
/// Erase Separate Files ///  
erase "wave1.dta"  
erase "wave4.dta"  
erase "wave5.dta"  
erase "weights.dta"
```

```
/****** STEP 2: RECODE VARIABLES *****/  
use "wave145.dta", clear  
recode y5 // Wave 5 outcome  
recode y4 // Wave 4 outcome  
recode x1 // Wave 1 predictor  
recode c4 // Wave 4 covariate  
save "wave145.dta", replace
```

```
/****** STEP 3: ACCOUNT FOR SURVEY DESIGN *****/  
svyset psuscid [pw = gsw145], strata(region)
```

```
/****** STEP 4: RUN MODEL *****/
```

```
svy: sem (y4 <- x1 c4) ///  
        (y5 <- y4 x1 c4)
```

```
/****** ALTERNATIVE SPECIFICATION *****/
```

```
sem (y4 <- x1 c4) ///  
    (y5 <- y4 x1 c4) ///  
    [pw=gsw145], vce(cluster psuscid)
```



Survey: Structural equation model  
 Number of strata = 4  
 Number of PSUs = 132  
 Number of obs = 10,640  
 Population size = 21,570,097  
 Design df = 128

	Coefficient	Linearized std. err.	t	P> t	[95% conf. interval]	
<b>Structural</b>						
<b>y4</b>						
x1	-.0216725	.0061663	-3.51	0.001	-.0338735	-.0094714
c4	.0347016	.0018049	19.23	0.000	.0311303	.0382729
_cons	.5092699	.0104647	48.67	0.000	.4885638	.529976
<b>y5</b>						
y4	.4598734	.013069	35.19	0.000	.4340142	.4857327
x1	.0257302	.0051315	5.01	0.000	.0155767	.0358837
c4	.0224808	.0018509	12.15	0.000	.0188184	.0261432
_cons	.1909264	.0123687	15.44	0.000	.1664528	.2154

Structural equation model  
 Estimation method: ml  
 Number of obs = 10,640

Log pseudolikelihood = -52826188  
 (Std. err. adjusted for 132 clusters in psuscid)

	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
<b>Structural</b>						
<b>y4</b>						
x1	-.0216725	.0061157	-3.54	0.000	-.0336589	-.009686
c4	.0347016	.0018158	19.11	0.000	.0311427	.0382605
_cons	.5092699	.0104592	48.69	0.000	.4887702	.5297696
<b>y5</b>						
y4	.4598734	.0136005	33.81	0.000	.433217	.4865299
x1	.0257302	.005109	5.04	0.000	.0157168	.0357436
c4	.0224808	.0018743	11.99	0.000	.0188073	.0261543
_cons	.1909264	.0125677	15.19	0.000	.1662941	.2155587

# Add Health Contact Info For Help with Data & Analytic Questions:

[addhealth@unc.edu](mailto:addhealth@unc.edu)

# Acknowledgements

Wave VI of Add Health is supported by two grants from the National Institute on Aging (1U01AG071448, principal investigator Robert A. Hummer, and 1U01AG071450, principal investigators Robert A. Hummer and Allison E. Aiello) to the University of North Carolina at Chapel Hill. Co-funding for Wave VI is being provided by the Eunice Kennedy Shriver National Institute of Child Health and Human Development, the National Institute on Minority Health and Health Disparities, the National Institute on Drug Abuse, the NIH Office of Behavioral and Social Science Research, and the NIH Office of Disease Prevention. Waves I-V data are from the Add Health Program Project, grant P01 HD31921 (Kathleen Mullan Harris) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. The content of this presentation is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health or the University of North Carolina at Chapel Hill.

Add Health was originally designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill. Add Health is currently directed by Robert A. Hummer; it was previously directed by Kathleen Mullan Harris (2004-2021) and J. Richard Udry (1994-2004).

Information on obtaining Add Health data is available on the project website (<https://addhealth.cpc.unc.edu>).

# Examples of Sampling Weights for Wave III Special Sub-Samples: Estimating Single-Level (population average) Models

Data Set (Year collected)	Sampling Weight Variable (N)	Sample	Represented Population of Interest
Wave III (2001)	W3PTNR (N=1,317)	Wave III Romantic Partner Sample: Eligible Wave I respondents and romantic partners interviewed at Wave III.	Romantic Partners of Adolescents enrolled in Grades 7-12 in 1994-1995
	TWGT3_2 (N=11,637)	Wave III Education Sample: Eligible Wave I respondents interviewed at Wave III.	Adolescents enrolled in Grades 7-12 in 1994-1995 who participated in high school transcript study
	TWGT3 (N=8,847)	Wave III Education Sample: Eligible Wave II respondents interviewed at Wave III.	Same as above

# Sampling Weights for Time-to-Event (Survival/Hazard) Analysis: One Time Point

Data Source (Y from One Wave)	Weight for Population Average Models	Weights for Multilevel Models	Sample	Target Population
Wave I only (1995)	GSWGT1 (N=18,924)	SCHWT1 (N=132) W1_WC (N=18,924)	Adolescents chosen with a known probability of being selected from 1994-1995 enrollment rosters of US schools.	Adolescents in 1995 enrolled in grades 7-12 during 1994-1995
Wave II only (1996)	GSWGT2 (N=13,570)	W2_WC (N=13,568)	Wave I respondents who were interviewed at Wave II.	Adolescents in 1996 enrolled in grades 7-12 during 1994-1995
Wave III only (2001)	GSWGT3_2 (N=14,322)	W3_2_WC (N=14,322)	Wave I respondents who were interviewed at Wave III.	Adults in 2001 enrolled in grades 7-12 during 1994-1995
Wave IV only (2008)	GSWGT4_2 (N=14,800)	W4_2_WC (N=14,800)	Wave I respondents who were interviewed at Wave IV.	Adults in 2008 enrolled in grades 7-12 during 1994-1995
Wave V only (2016-18)	GSWGT (N=12,300)	W5_2_WC (N=12,300)	Wave I respondents who were interviewed at Wave V.	Adults in 2016-18 enrolled in grades 7-12 during 1994-1995

# Sampling Weights for Time-to-Event (Survival/Hazard) Analysis: Multiple Time Points

Data Source (Y from Multiple Waves)	Weight for Population Average Models	Weights for Multilevel Models	Target Population
Wave I & II	GSWGT1 (N=18,924)	SCHWT1 (N=132) W1_WC (N=18,924)	Adolescents in 1995 enrolled in grades 7-12 during 1994-1995
Wave II & III	GSWGT2 (N=13,570)	SCHWT1 (N=132) W2_WC (N=13,568)	Adolescents in 1996 enrolled in grades 7-12 during 1994-1995
Wave I, II, & III	GSWGT1 (N=18,924)	SCHWT1 (N=132) W1_WC (N=18,924)	Adolescents in 1995 enrolled in grades 7-12 during 1994-1995
Wave I, II, III, & IV	GSWGT1 (N=18,924)	SCHWT1 (N=132) W1_WC (N=18,924)	Adolescents in 1995 enrolled in grades 7-12 during 1994-1995
Wave I, II, III, IV & V	GSWGT1 (N=18,924)	SCHWT1 (N=132) W1_WC (N=18,924)	Adolescents in 1995 enrolled in grades 7-12 during 1994-1995