Introduction to the National Longitudinal Study of Adolescent to Adult Health (Add Health)

Hsueh-Sheng Wu
Center for Family and Demographic Research
January 23, 2023
Outline

• Introduction
• What is special about Add Health?
• Survey design
• Subject areas
• Add Health Codebook Explorer
• Data files
• File location
• Unit of analysis
• Analytic tips
• Studies using Add Health data
• Help with Add Health analyses
• Conclusions
Introduction

• National Longitudinal Study of Adolescent Health (Add Health) is a study that the Carolina Population Center at the University of North Carolina-Chapel Hill has conducted to follow a nationally representative sample of adolescents in grades 7-12 since 1994.

• These adolescents were first interviewed in 1994-1995 (Wave I) and followed up in 1996 (Wave II), 2001-2002 (Wave III), 2007-2008 (Wave IV) and finally 2016-2019 (Wave V).

• Add Health also has supplemental education data. In fall 2001, the Population Research Center at the University of Texas-Austin collected data that supplement Add Health. These supplementary data focus on (1) educational achievement, (2) course taking patterns, (3) curricular exposure, and (4) educational contexts of Add Health respondents at Wave III.

• In addition to original Add Health study, three satellite projects have been initiated, including (1) the Add Health Parent Study (Parents 2015-2017), (2) Sexual Orientation/Gender Identity, Socioeconomic Status, and Health across the Life Course (SOGI-SES)”, and (3) the Omics project that examines the relations between biological factors (i.e., gene, proteins, molecules) and various outcomes.
What Is Special about Add Health?

• Add Health is an ongoing study that has the largest representative sample of adolescents and collects the most comprehensive data of these adolescents for over 20 years.

• Add Health started with 20,745 adolescents at wave 1 (1994-1995) and more than 20 years later, it still have 12,300 adolescents at Wave V (2016-2019).

• Add Health collected comprehensive information on biological and psychological developments of adolescents and the social contexts, such as home, friends, intimate relationships, schools, and neighborhoods.

• Because of its sample size, scopes of data, and panel design, Add Health is invaluable for examining how social, psychological, and biological factors influence an individual’s life through adolescence, young adulthood, and later adulthood.
Survey Design

• Add Health used a stratified two-stage sampling method:
  – The sampling frame is stratified by region, urbanization, school size, school type, and race composition
  – 80 high schools and 52 middle schools were selected with an unequal probability at the first stage
  – 90,000 students were selected to fill out in-school Add Health questionnaire, and 20,745 of them fill out in-home questionnaire

• Add Health oversampled twins and siblings of twins; non-related adolescents residing together; disabled minority students; Blacks from well-educated families; and minority students who are Chinese, Cubans, and Puerto Ricans.

• Data were collected with Computer-Assisted Personal Interviewing (CAPI) and questionnaire.
Survey Design (Cont.)

Supplemental Education Data:

- The sample consists of all respondents at Wave III of Add Health.
- The study collected high school transcripts and other data from high schools that Add Health respondents last attended.
- The data were collected from 130 Add Health high schools and 1,400 additional high schools.
- Education data were collected for approximately 12,000 respondents, which is about 80% of Add Health respondents at Wave III.
### Tab 1. The Life Segments covered by Wave V of Add Health (N=12,300)

| Age at W1 | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  | 21  | 22  | 23  | 24  | 25  | 26  | 27  | 28  | 29  | 30  | 31  | 32  | 33  | 34  | 35  | 36  | 37  | 38  | 39  | 40  | 41  | 42  | 43  | 44  | 45-64 | Total |
|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 11        | 1   | 2   | 2   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 6    |
| 12        |     |     | 1,2 | 2   | 2   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 321  |
| 13        |     |     |     |     | 1,2 | 2   |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 1,421|
| 14        |     |     |     |     |     |     | 1,2 |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 1,735|
| 15        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 2,152|
| 16        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 2,362 |
| 17        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 2,300 |
| 18        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 1,693 |
| 19        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 262   |
| 20        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 41    |
| 21        |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     | 7     |

**Note:** The table above details the life segments covered by Wave V of Add Health (N=12,300) across different age groups. The table shows the distribution of respondents across various age groups from adolescence to midlife.
Subject Areas

• Add Health covers many interesting subject areas.

• Some of the areas have been covered at each wave, whereas others are covered at only certain wave or waves.

• The list of sections in the codebooks at each wave shows what subject areas are covered at each wave.

• When a subject area is covered across waves, it does not necessarily mean all questions about the subject area are asked across waves. So, it is very important to check at which waves a certain question is asked.
Add Health Codebook Explorer (ACE)

Welcome to the Add Health Codebook Explorer (ACE). This tool allows you to view questions asked of respondents, parents, and interviewers during the In-School and In-Home interviews in Waves I through IV.

You can browse by topic or search for questions by variable name, keyword, or phrase. 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. The questions included in ACE are from the In-School and In-Home interviews. There are many other data sets available; please visit the Codebooks page for a complete list of available data, including contextual and biological data.

Please refer to the ACE instructions before searching and browsing to maximize your results.
Add Health Codebook Explorer (Cont.)
Data Files

• CFDR stores a copy of public data in the public folder (R:\CFDR\Public\Data\AddHealth). In addition, public data can be downloaded from ICPSR website (http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/21600).

• CFDR stores a copy of restricted data on the secured server (R:\AddHealth). Only people who have obtained the permission from the Carolina Population Center at the University of North Carolina-Chapel Hill can access the data.

• The difference between the public data and the restricted data is that public data contain about only one-third of respondents, whereas the restricted data have all respondents.
• Supplemental Education Data
  – All education data are restricted data
  – CFDR stores a copy of restricted data on the secured server (\AddHealth). Only people who have obtained the permission from the Carolina Population Center at the University of North Carolina-Chapel Hill can access the data

• CFDR has constructed some SAS data sets from the restricted Add Health data, including the in-home interview data, weighted data, and family structure measures from Wave I through IV. These constructed data are stored in the folder “\AddHealth\ADD Health\Add Health study\CFDR SAS data.”

• If you need to use restricted Add Health data, please contact Dr. Jenjira Yahirun(yahirun@bgsu.edu).
File Location

Folder name: R:\
File Location (Cont.)

Former name: R:\AddHealth
File Location (Cont.)
Folder name: R:\AddHealth\ADD Health
File Location (Cont.)

Folder name: R:\AddHealth\ADD Health\Add Health study

<table>
<thead>
<tr>
<th>Name</th>
<th>Date Modified</th>
<th>Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>CFDR SAS data</td>
<td>7/12/2020 8:22 PM</td>
<td>File folder</td>
<td></td>
</tr>
<tr>
<td>data and related documents</td>
<td>5/6/2020 2:31 PM</td>
<td>File folder</td>
<td></td>
</tr>
<tr>
<td>general information</td>
<td>11/26/2018 11:21 PM</td>
<td>File folder</td>
<td></td>
</tr>
<tr>
<td>tips for analyzing Add Health data</td>
<td>11/26/2018 11:21 PM</td>
<td>File folder</td>
<td></td>
</tr>
</tbody>
</table>
Main Add Health Data Folders: “R:\AddHealth\ADD Health\Add Health study\data and related documents”
File Location (Cont.)

The location of first four waves of In-home interview data: “R:\AddHealth\ADD Health\Add Health study\data and related documents\in-home interview at Waves I, II, III, and IV”
The location of In-home interview data at Wave V: “R:\AddHealth\ADD Health\Add Health study\data and related documents\in-home interview at Waves I, II, III, and IV\in-home interview at Wave V”
The location of In-home interview data at Wave V: “R:\AddHealth\ADD Health\Add Health study\data and related documents\in-home interview at Waves I, II, III, and IV\in-home interview at Wave V\data”
CFDR puts our constructed data and variables in a specific folder: R:\AddHealth\ADD Health\Add Health study\CFDR SAS data
File Location (Cont.)
The content of the folder: R:\AddHealth\ADD Health\Add Health study\CFDR SAS data\Constructed Data
All Five Waves of In-Home interview individual data are in the folder “R:\AddHealth\ADD Health\Add Health study\CFDR SAS data”
Unit of Analysis

- Add Health collects information on individual adolescents, their social environment (e.g., neighborhood, school, family) and various aspects of social relations and experiences (e.g., intimate relationship, pregnancy, live births, and parent-children relationship).

- An example of the nested structure of neighborhood, school, family, and individual adolescents is as follows:
An example of the nested structure of Individual, relationship, pregnancy, live births, and parent-children relationship is as follows:

- Student 1
  - Relationship 1
    - Pregnancy
      - Live-birth
        - Child 1
  - Relationship 2
    - Pregnancy
      - Live-birth
        - Child 2
      - Live-birth
        - Child 3
Analytic Tips

• How to find the variables you need?
• How to read Add Health data?
• How to merge data?
• How to weight Add Health data?
• How to change the unit of analysis?
How to Find the Variables You Need?

• Use codebooks or Add Health Codebook Explorer to locate the variables of interest

• Add Health provides codebooks that list all of the names and wordings of the variables at each wave. Thus, if you are interested in the in-home interview data, you should start finding your variables by reading through the following codebooks:
  – WAVE1NDX.PDF
  – WAVE2NDX.PDF
  – wave3ndx.pdf
  – wave4ndx.pdf
  – WaveVMixed_ModeSurveyCodebook.pdf and WaveVSection16BCodebook

  – Each subject area usually has its own codebook, and you can only find value labels in each codebook
How to Read Add Health Data?

• The public data of Add Health may be in SAS, Stata, or SPSS format. You can use Stat/Transfer to change the data from one format to another.

• The restricted data of Add Health are initially in SAS export format. The following codes provide instructions on how to use SAS and Stata to read in the SAS export file.
How to Read Add Health Data? (Cont.)

SAS code:

LIBNAME wave1 xport “R:\ADD Health\Add Health study\data and related documents\in-home interview at Waves I, II, and III\in-home interview at Wave I\data\allwave1.exp“;

LIBNAME out “r:\Add Health\temp”;

Data out.wave1;
Set wave1.allwave1;
Run;

Stata code:

fdause “R:\ADD Health\Add Health study\data and related documents\in-home interview at Waves I, II, and III\in-home interview at Wave I\data\allwave1.exp“
describe
How to Merge Data?

• When do data need to be merged?
  – If you want to combine data from different waves of Add Health
  – If you want to combine data with different unit of measurements
  – If you want to use both Add Health data and Education data

• SAS and Stata sample commands to merge Waves I and II data are shown in the following slides:
How to Merge Data? (Cont.)

- SAS code:

```sas
Libname in "R:\AddHealth";
************************;
PROC SORT DATA=in.wave1;
   BY aid;
RUN;
************************;
PROC SORT DATA=in.wave2;
   BY aid;
RUN;
************************;
DATA in.wave12;
   MERGE in.wave1 (IN=in_wave1) in.wave2 (IN=in_wave2);
   BY aid;
RUN;
```
How to Merge Data? (Cont.)

• Stata code:

```
use "R:\AddHealth\wave1.dta"
sort aid
save "R:\AddHealth\wave1_2.dta", replace

use "R:\AddHealth\wave2.dta"
sort aid
save "R:\AddHealth\wave2_2.dta", replace

use "R:\AddHealth\wave1_2.dta", clear
sort aid
merge aid using "R:\AddHealth\wave2_2.dta"
tab1 _merge
rename _merge wave12
label variable wave12 "indicator for merging waves 1 and 2"
sort aid
save "R:\AddHealth\wave12.dta", replace
```
How to Weight Add Health Data?

• Add Health data are collected with a complex survey design. Therefore, each respondent does not have the same probability of being selected into the sample and thus needs to be reweighted.

• Clustering of students from the same regions and schools.

• The analysis of Add Health data always needs to be weighted in order to adjust for the effects of its complex survey design.

• SAS and Stata differ in their abilities of performing statistical analyses, while controlling for the effects of the complex survey design.
### How to Weight Add Health Data? (Cont.)

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Stata command</th>
<th>SAS command</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate means for survey data</td>
<td>svy: mean</td>
<td>Proc Surverymeans</td>
</tr>
<tr>
<td>Estimate proportions for survey data</td>
<td>svy: tab</td>
<td>Proc Surveyfreq</td>
</tr>
<tr>
<td>Linear regression for survey data</td>
<td>svy: regress</td>
<td>Proc Surveyreg</td>
</tr>
<tr>
<td>Logistic regression for survey data, reporting odds ratios</td>
<td>svy: logistic</td>
<td>Proc Surveylogistic</td>
</tr>
<tr>
<td>Cox proportional hazards model for survey data</td>
<td>svy: stcox</td>
<td>Proc Surveyphreg</td>
</tr>
<tr>
<td>Ordered logistic regression for survey data</td>
<td>svy: ologit</td>
<td></td>
</tr>
<tr>
<td>Ordered probit regression for survey data</td>
<td>svy: oprobit</td>
<td></td>
</tr>
<tr>
<td>Multinomial (polytomous) logistic regression for survey data</td>
<td>svy: mlogit</td>
<td></td>
</tr>
<tr>
<td>Multinomial probit regression for survey data</td>
<td>svy: mprobit</td>
<td></td>
</tr>
<tr>
<td>Parametric survival models for survey data</td>
<td>svy: streg</td>
<td></td>
</tr>
<tr>
<td>Generalized linear models for survey data</td>
<td>svy: glm</td>
<td></td>
</tr>
<tr>
<td>Generalized negative binomial regression for survey data</td>
<td>svy: gnbreg</td>
<td></td>
</tr>
<tr>
<td>Poisson regression for survey data</td>
<td>svy: poisson</td>
<td></td>
</tr>
<tr>
<td>Zero-inflated negative binomial regression for survey data</td>
<td>svy: zinb</td>
<td></td>
</tr>
<tr>
<td>Zero-inflated Poisson regression for survey data</td>
<td>svy: zip</td>
<td></td>
</tr>
</tbody>
</table>
How to Weight Add Health Data? (Cont.)

If you use the full sample in the analysis, you can use either SAS or Stata for the analysis.

**SAS code:**

```sas
Libname in "R:\AddHealth\TEMP";

proc surveylogistic data= in.logit3;
cluster psuscid3;
weight gswgt3;
strata region3;
model h3ed3 = bio_sex3 calcage3;
run;
```

**Stata code:**

```stata
use "R:\AddHealth\TEMP\logit3.dta", clear
svyset psuscid3 [pweight =gswgt3], strata(region3)
svy: logit h3ed3 bio_sex3 calcage3
```
How to Weight Add Health Data? (Cont.)

If you use only part of the sample in the analysis, you should use Stata for the analysis because SAS has the sub-population options for the Proc Surveymeans command only.

Stata code:

```stata
use "R:\AddHealth\TEMP\logit3.dta", clear
svyset psuscid3 [pweight =gswgt3], strata(region3)
svy, subpop(marker): logit h3ed3 bio_sex3 calcage3
```
How to Change the Unit of Analysis?

- Changing the unit of analysis means changing the unit of observations in the data set. Because of the nested structure of Add Health data, you can change the unit of observations from one level to another.

- When the unit of analysis changes, the number of valid observation changes, too.

<table>
<thead>
<tr>
<th>Unit of Analysis</th>
<th>Number of Analysis Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHILD</td>
<td>4,181</td>
</tr>
<tr>
<td>BIRTH</td>
<td>4,181</td>
</tr>
<tr>
<td>PREGNANCY</td>
<td>4,055</td>
</tr>
<tr>
<td>RELATION</td>
<td>3,293</td>
</tr>
<tr>
<td>RESPONDENT</td>
<td>2,960</td>
</tr>
</tbody>
</table>

- SAS and Stata examples of changing unit from birth to pregnancy.
How to Change the Unit of Analysis? (Cont.)

- SAS code:

```
PROC SORT DATA = out.sect25 OUT=out.preg;
BY aid rrelno rpregnno;
RUN;

PROC FREQ DATA = out.preg;
TABLES birthno;
RUN;

PROC TRANSPOSE DATA = out.preg
OUT=out.f_preg PREFIX =birth;
BY aid rrelno rpregnno;
ID birthno;
VAR c_age;
RUN;

PROC CONTENTS DATA = out.sect25;
RUN;

PROC CONTENTS DATA = out.f_preg;
RUN;
```
How to Change the Unit of Analysis? (Cont.)

- Stata code:

```
use t:\temp\sect25.dta, clear

des aid rrelno rpregno

sum rrelno rpregno

tostring rrelno, generate(srrelno)
tostring rpregno, generate(srpregno)

gen said_rp3 = aid + srrelno + srpregno
replace said_rp3 = aid + "0" + srrelno + srpregno if rrelno > =1 & rrelno <=9

label variable said_rp3 "string id for pregnancy record"

sort said_rp3

des

tab1 birthno

reshape wide c_age, i(said_rp3) j(birthno)

des

rename c_age1 birth1
rename c_age2 birth2

save "t:\temp\pregnancy.dta", replace
```
Studies Using Add Health Data

• There have been more than 4,000 publications using Add Health. You can locate them through Add Health or ICPSR web site:
  – Add Health Web site
    • https://addhealth.cpc.unc.edu/publications/
  – ICPSR website:
    • After you find Add Health data (https://www.icpsr.umich.edu/web/ICPSR/series/1006/studies), click on the tab “Data-related Publications”
Help with Add Health Analyses

• **Official Add Health listserv**
  Listserve is a place where Add Health users ask and answer questions about analyzing Add Health data. To subscribe the Add Health listserv, send e-mail to: listserv@unc.edu and in the body of the message put: subscribe addhealth2 <firstname lastname>

• **Add Health Users Conference**
  Carolina Population Center at University of North Carolina has hosted 14 Add Health Users Conferences on how to construct and analyze Add Health data. Add Health website ([http://www.cpc.unc.edu/projects/addhealth/news](http://www.cpc.unc.edu/projects/addhealth/news)) provides information on the upcoming Add Health User Conference.

• **CFDR Programming Help**
  If you have programming problems, contact Hsueh-Sheng Wu at whu@bgsu.edu
Conclusions

• Add Health is an excellent data set for studying how adolescents make transitions into adulthood and how such transitions are influenced by their families, peers, schools, and neighborhoods.

• When using Add Health data, researchers should allow themselves sufficient time to construct variables because they may need to correct for possible inconsistencies in data from different measurement units and/or across different waves before they can construct variables for their analyses.

• The analysis of Add Health data always needs to be weighted in order to adjust for its complex survey design. This workshop covers how to weight the results if the single-level analysis is conducted. If multilevel analysis is to be conducted, researchers need to rescale the weights, so units at different levels have their respective weights. For details, consult the document, Guidelines for Analyzing Add Health data (https://addhealth.cpc.unc.edu/wp-content/uploads/docs/user_guides/GuidelinesforAnalysisofAddHealthData_202004.pdf)

• If you have any questions about using, coding, or analyzing Add Health data, please feel free to contact me (wuh@bgsu.edu).