

# Overview of the National Longitudinal Survey of Youth 1997 (NLSY97)

Hsueh-Sheng Wu  
CFDR Workshop Series  
November 9, 2015

BGSU



Center for  
**Family and  
Demographic** Research

# Outline

- Introduction to the NLSY97
- Advantages and challenges of using NLSY97
- How to use NLS Investigator website
- How to link data from different rounds together
  - Linking roster Items across rounds
  - Finding the first partner's age
- How to use weight variables
- Journal articles using NLSY 97
- Sample Stata codes for two types of linking in NLSY97
- Conclusions

# Introduction to the NLSY 97

- NLSY97 is one of the five cohort studies conducted by the U.S. Department of Labor's Bureau of Labor Statistics.
- NLSY97 initially was designed to document how the youths transit from school to work and into adulthood.
- The Subject Areas of NLSY97 now include:

Labor market behavior	Relationships with parents
Educational experiences	Contact with absent parents
Training	Marital and fertility histories
Participation in government assistance programs	Dating
Expectations	Sexual activity
Criminal behavior	Onset of puberty
Alcohol and drug use	Time use

# Introduction to the NLSY 97 (Continued)

- NLSY 97 is a cohort survey whose target cohort is youths who were 12 to 16 years old as of December 31, 1996.
- NLSY97 was conducted annually between 1997 and 2012 and biennially afterwards. Currently, sixteen rounds of data were available for public use through NLS Investigator website. Not every subject topic was examined in every round of NLSY97.
- NLSY97 used a multi-stage stratified sampling method to select eligible households. Blacks and Hispanic were over-sampled. At the first phase, 96,512 households were selected from 1,748 sample segments embedded within 147 primary sampling units (PSUs). At the second phase, screening interviews were used to identify 9,907 households with eligible youths, and 8,984 of them participated in the 1997 survey.
- The NLSY97 data were mainly collected using Computer-Assisted Personal Interview(CAPI). Audio Computer-Assisted Self-Interview (ACASI), and questionnaires,

# Advantages and Challenges of Using NLSY97

## Advantages:

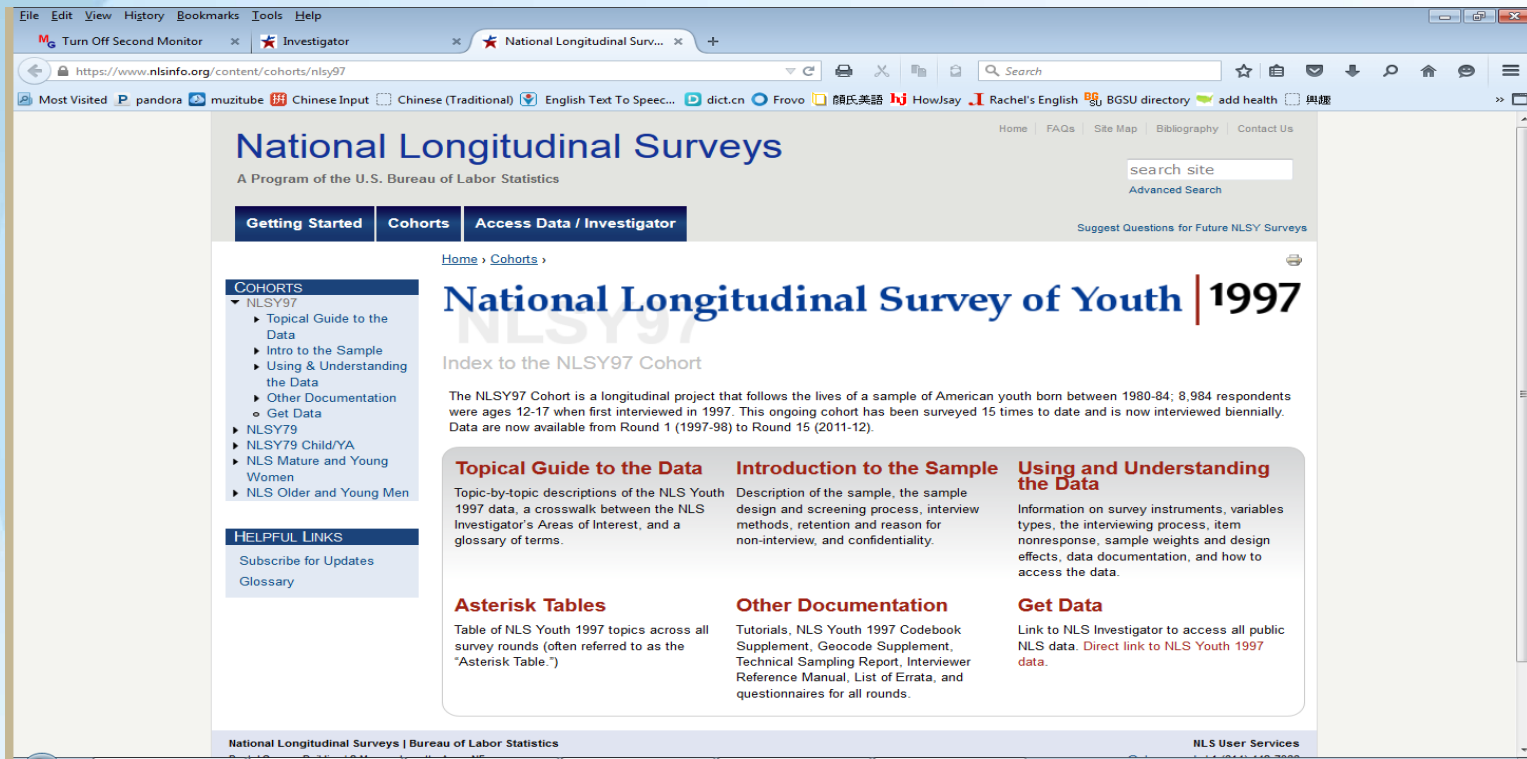
- A variety of subject areas covered in NLSY97.
- A clearer temporal order of variables.
- Individual heterogeneity is better controlled.
- A better understanding of how youths transitioned into adulthood and what may influence such a transition.
- Many roster variables have been created for research use.
- Provide custom weight variables

## Challenges:

- Generalizability problem (Findings are cohort-specific.)
- Changes to designs and questionnaire of NLSY97 over time
- The original raw data are too large to be directly downloaded and analyzed. Thus, researchers need to learn how to use NLS Investigator website to download the data they need.
- The availability of 16 rounds of data creates the problems of linking and constructing variables across rounds of surveys.
- Different analytic strategies are needed if more than two rounds of data are used.

# How to Use NLS Investigator Website

- NLSY97 main webpage (<https://www.nlsinfo.org/content/cohorts/nlsy97>)



- NLSY97 User's Guide (Rounds 1-13)
- <https://www.nlsinfo.org/site/nlsy97/nlsdocs/nlsy97/mainto>

# How to Use NLS Investigator Website

- NLS Investigator

<https://www.nlsinfo.org/investigator/pages/login.jsp>

The screenshot displays the NLS Investigator website interface. At the top, there is a navigation bar with the title "NLS Investigator" and a welcome message "Welcome, Hsueh-Sheng". The main content area features a dropdown menu for "Select the study you want to work with:" set to "NLSY97 1997-2013 (rounds 1-16)", with a release date of "Released September 25, 2015". To the right, under "Additional Resources:", there are links for "Errata, Documentation" and "Custom Weights". A button labeled "To start a new search click here" is also present. Below this, a series of tabs includes "Choose Tagsets", "Variable Search", "Review Selected Variables (6)", "Codebook", and "Save / Download". The "Variable Search" tab is active, showing a "Browse Index" section with a list of categories and their respective variable counts. A pop-up window is open over the "Options" dropdown, displaying the text: "Please browse the index on the left to display variables. This index contains a set of NLSY97 variables commonly used in research and is not the full data set." The footer contains links for "NLS Home", "NLS Bibliography", and "Privacy Policy", along with contact information for usersvc@chrr.osu.edu.

File Edit View History Bookmarks Tools Help

Turn Off Second Monitor Investigator

https://www.nlsinfo.org/investigator/pages/search.jsp

Most Visited pandora muzitube Chinese Input Chinese (Traditional) English Text To Speec... dict.cn Frovo 顏氏美語 HowSay Rachel's English BGSU directory add health 興趣

# NLS Investigator

Welcome, Hsueh-Sheng | [SEARCH](#) | [Downloads](#) | [Account](#) | [Help](#) | [Logout](#)

Select the study you want to work with:  
NLSY97 1997-2013 (rounds 1-16)  
Released September 25, 2015

Additional Resources:  
[Errata, Documentation](#) (user's guide, questionnaires and other materials)  
[Custom Weights](#)

To start a new search [click here](#)

Choose Tagsets Variable Search Review Selected Variables (6) Codebook Save / Download

Browse Index Browse Index with Search Search

Index of Selected Variables Options

Education, Training & Achievement Scores (21640)  
Employment (24490)  
Household, Geography & Contextual Variables (8051)  
Dating, Marriage & Cohabitation (1467)  
Sexual Activity, Pregnancy & Fertility (1601)  
Children (1909)  
Parents, Family Process & Childhood (175)  
Income, Assets & Program Participation (4105)  
Health (1643)  
Attitudes, Expectations & Non-cognitive Tests (422)  
Crime & Substance Use (6734)  
Survey Methodology (354)

Please browse the index on the left to display variables.  
This index contains a set of NLSY97 variables commonly used in research and is not the full data set.

NLS Home | NLS Bibliography | Privacy Policy

For help, email [usersvc@chrr.osu.edu](mailto:usersvc@chrr.osu.edu)

# How to Use NLS Investigator Website

## Obtaining NLSY 97 data:

Step 1. Register with the website

Step 2. Browse or search variables

Step 3. Select the variables you want

Step 4. Download the data

Step 5. Run the corresponding SAS, SPSS, or Stata command codes to create the data file for your analyses



# How to Link Data from Different Rounds together

- Linking data from different rounds is necessary to describe how individuals remain the same or change over time.
- Use NLSY97 for two types of linking:
  - The first type is to link the characteristics of a respondent over time. For example, researchers may want to understand how the respondent's marital status, educational attainment, or employment status changed over time.
  - The second type is to link the characteristics of people related with respondents over time. For example, researchers may want to understand the gender and education attainment of all spouses that an respondent have had.
  - The second type of linking is more difficult than the first type of linking.
- NLSY website provides sample SAS codes for these two types of linking.
  - <https://www.nlsinfo.org/content/getting-started/intro-to-the-nls/tutorials/tutorial-linking-roster-items-across-rounds>
  - <https://www.nlsinfo.org/content/getting-started/intro-to-the-nls/tutorials/tutorial-matching-cohabiting-partners-their>

# How to Link Data from Different Rounds Together

Table 1. Example of Linking Marital Status across Rounds in NLSY 97

Respondents	Rounds					
	1	2	3	4	5	6
Bob	never married	never married	never married	never married	never married	never married
Jim	never married	never married	never married	never married	married	divorced
Kim	never married	never married	married	married	married	married
Matt	never married	married	divored	married	divored	married

Table 2. Example of Different Spouses at Rounds of NLSY 97

Respondents	Rounds					
	1	2	3	4	5	6
Bob	never married	never married	never married	never married	never married	never married
Jim	never married	never married	never married	never married	married to Sue	divorced
Kim	never married	never married	married to Terry	married	married	married
Matt	never married	married to Pam	divored	married to Ann	divored	married to Jesey

# How to Link Data from Different Rounds Together

The complexities involved:

- The marital history roster indicates whether a respondent has a spouse or partner at a time, which allows for identifying when a respondent was married or cohabiting. However, this roster provides the position, rather than the unique identification number (i.e. the ID variable) of the spouses or partners.
- Researchers need to use the position that the spouse or partner hold at the round of survey to find the unique ID assigned to the spouse or partner at the time.
- NLSY97 stores the information of spouses/partners in three different files: the marriage and cohabitation section, the household roster, and the non-resident roster. These files differ in how to extract the information about spouses/partners.
- The members in the marriage and cohabitation section, the household roster, and the non-resident roster change over time and require different ways of merging with other files.

# How to Link Data from Different Rounds Together

Five steps of conducting the second type of linking:

- Determine when a respondent had his/her first partner and what position the partner holds in the partner roster at that time. This will be the master data file.
- Construct separate data files from the marriage and cohabitation section, the household roster, and the non-resident roster.
- Merge the master data file with the data from the marriage and cohabitation section using the partner position at that time.
- Merge the master data file with the data from the household roster using the partner ID.
- Merge the master data file with the non-resident roster using the partner ID.

# How to Use Weight Variables

- NLSY97 weight variables can be customized, depending on how many rounds of surveys are used and whether the focus is on respondents participating in all or any of these rounds.
- The custom weight variables were created by adjusting for differences in base weight, nonresponse to the screener, different response rates to the screener, different cross-sectional weights for Black and Hispanic respondents, nonresponse to the main interview.
- NLSY97 custom weight variable is a post-stratification person weight variable.
- NLSY97 custom weight variables can be downloaded from the website:  
<https://www.nlsinfo.org/content/cohorts/nlsy97/using-and-understanding-the-data/sample-weights-design-effects>

# Journal Articles Using NLSY 97

Manning, Wendy D. (2015) Family Formation Processes: Assessing the Need for a New Nationally Representative Household Panel Survey in the United States. *Journal of Economic and Social Measurement*, 40(1-4), 197-219.

Guzzo, K.B. (2015). Young parents and co-residence with their own parents (2015 Working Paper Series). Retrieved from California Center for Population Research: <http://papers.ccpr.ucla.edu/papers/PWP-BGSU-2015-008/PWP-BGSU-2015-008.pdf>

Sandberg-Thoma, S.S.; Snyder, A.R. & Jang, B.J. (2015). Exiting and Returning to the Parental Home for Boomerang Kids, *Journal of Marriage and Family*, 77(3), 806-818. <http://DX.DOI: 10.1111/jomf.12183>

Tumin, D.; Siqi Han, S. & Qian, Z. (2014). Estimates and Meanings of Marital Separation. *Journal of Marriage and Family*, 77(1), 312-322. <http://DX.DOI: 10.1111/jomf.12149>

# Conclusions

- NLSY97 provided an excellent opportunity to examine how the youth transitioned into the adulthood. With measures of various subject areas, NLSY may shed light on why some youths made the transition while some did not.
- You need to decide your analytic models first before linking and constructing variables.
- Given the size and complexity of NLSY97, please remember that it will take lots of time to link or construct variables using NLSY97.
- Both SAS and Stata can link and construct the variables of interest. If you like to work with data in the wide format, you can use SAS. If you like to work with data in the long format, you can use Stata.
- If you have any questions about using the NLSY97 data. Please feel free to contact Hsueh-Sheng Wu at 372-3119 or [wuh@bgsu.edu](mailto:wuh@bgsu.edu)