

NLS DATA SETS AND DOCUMENTATION

National Longitudinal Surveys (NLS) data are available on cross-cohort CDs and as downloadable files on the Internet. Each data set contains data and documentation files, as well as software that allows users to peruse the data, select variables of interest, and create extract files. Electronic and paper-copy documentation items accompany each dataset and aid in its use.

This chapter first describes the current data releases and features of the search and extraction software. Following this discussion, readers will find descriptions of the various documentation items that will assist them in their work. Electronic and paper documentation items for the NLSY97, the NLSY79, the NLSY79 children and young adults, and the original cohorts are reviewed in separate subsections within this chapter.

NLS User Services Office

The items discussed in this chapter are available through NLS User Services. Individuals wishing to obtain NLS data and documentation can check pricing information, download data and documentation, and place orders through the www.bls.gov/nls Web site. Chapter 10 describes additional publications and research reports. Questions or problems can be addressed to NLS User Services using the following information:

NLS User Services Office
(614)442-7366
usersvc@postoffice.chrr.ohio-state.edu

Note to individuals ordering geocode data. To protect the confidentiality of respondents, the Bureau of Labor Statistics (BLS) grants access to NLSY79, NLSY79 young adult, and NLSY97 geocode files only to researchers in the United States who agree in writing to adhere to the BLS confidentiality policy and whose projects further the mission of BLS to conduct sound, legitimate research in the social sciences. Researchers interested in learning more about the NLS geocode program or downloading the application should see the NLS Web site at www.bls.gov/nls/geocodeapp.htm. Some geographic information from interviews with the original cohorts is available for research use at the U.S. Census Bureau's Research Data Centers. Information about applying for access to these data also is available on the NLS Web site.

Note to customers outside the continental United States. If data sets or documents are not available or cannot be accessed electronically, cross-cohort CDs and paper copies can be ordered. To cover costs, NLS User Services must charge handling fees to individuals ordering NLS data or documentation to be shipped to an address outside of the continental United States. Upon receipt of an order, User Services will calculate the cost for the materials and the type of mail service (air or surface) specified on the order form and will notify the customer of the amount of the handling charges for that order. NLS materials will be shipped upon receipt, in U.S. dollars, of both the cost of the items ordered and the handling fee.

Current data releases

Figure 8.1 lists the NLS data sets currently available to the public. Researchers can download data files and accompanying documentation directly to a computer or server for free. For a fee, researchers can order data files on the cross-cohort CD, as well as paper versions of NLS documentation. To download free data and documentation files, view pricing information, or order cross-cohort CDs and paper documentation, see the NLS Web site at www.bls.gov/nls.

NLSY97 data

The National Longitudinal Survey of Youth 1997 (NLSY97) data are available to researchers in three data sets:

- *The NLSY97 main file* contains the data and documentation from survey rounds 1 through 7, together with search and extraction software. The data set holds the data elements—described in detail in chapter 2—from the youth questionnaires; the rounds 2 through 5 household income updates; the round 1 parent questionnaire; and the round 1 screener, household roster, and non-resident roster questionnaire. This data set includes a set of *ASVAB* scores and information from the high school transcript data collections.
- *The NLSY97 event history file, including main file data*, contains all of the data on the main file, as well as the search and extraction software. Also included are a number of created event history variables describing the youth's employment, education, marriage, and program

Figure 8.1. NLS data files

Cohort	Data Files	Survey years
<i>Available by Download</i>		
NLSY97	<ul style="list-style-type: none"> • NLSY97 main files • NLSY97 event history file with main files 	Rounds 1–7
NLSY79	<ul style="list-style-type: none"> • NLSY79 main files with work history data 	1979–2002
NLSY79 Children/ Young Adults	<ul style="list-style-type: none"> • NLSY79 child and young adult files 	1986–2002 (child) 1994–2002 (young adult)
Mature and Young Women	<ul style="list-style-type: none"> • NLS of Mature and Young Women 	1967–2003 (Mature Women) 1968–2003 (Young Women)
Older and Young Men	<ul style="list-style-type: none"> • NLS of Older Men and Young Men 	1966–90 (Older Men) 1966–81 (Young Men)
<i>Available on CD</i>		
Cross-cohort	<ul style="list-style-type: none"> • Current public release of all cohorts 	
NLSY97	<ul style="list-style-type: none"> • NLSY97 geocode file with main files, event history data 	Rounds 1-7
NLSY79	<ul style="list-style-type: none"> • NLSY79 geocode data with main files, work history data 	1979-2002
NLSY79 young adults	<ul style="list-style-type: none"> • NLSY79 young adult geocode data with child and young adult files 	1994-2002

participation through round 7. This data set includes a set of *ASVAB* scores and information from the high school transcript data collections.

- ***The NLSY97 geocode data CD, including main file and event history data***, contains all of the main file data and documentation, the created event history variables, and detailed geographic data. Because the geographic data are confidential, only individuals who complete the BLS geocode application process can obtain this file. Information on obtaining this file can be found at www.bls.gov/nls/geocodeapp.htm.

The NLSY97 main file and event history data are available on the cross-cohort CD or can be downloaded from the Internet; the geocode file is available only on individual cohort geocode CDs.

NLSY79 data

Data from the National Longitudinal Survey of Youth 1979 (NLSY79) are available in two data sets:

- ***The NLSY79 main file with work history data*** contains data and documentation, together with search and extraction software, for the NLSY79 main interviews from 1979 to 2002. It also contains created work history variables detailing respondents' employment since 1978. This data set holds all of the data elements described in

chapter 3 for the main NLSY79 interviews, except for the detailed NLSY79 geocode data and the kinship proximity variables.

- ***The NLSY79 geocode data CD with main file and work history data*** contains the constructed NLSY79 geocode files from 1979 to 2002, the NLSY79 main survey data for the same years, and the NLSY79 work histories. This CD holds all data elements described in chapter 3, except for the kinship proximity variables. Because the CD contains detailed confidential geographic data, it is available only to individuals who satisfactorily complete the BLS geocode application process. Information on obtaining this file can be found at www.bls.gov/nls/geocodeapp.htm.

The NLSY79 main file and work history data are available on the cross-cohort CD or can be downloaded from the Internet; the geocode file is available only on individual cohort geocode CDs. Researchers who are interested in the NLSY79 women's support network data file described in chapter 3 should contact NLS User Services (e-mail: usersvc@postoffice.chrr.ohio-state.edu).

NLSY79 child and young adult data

The National Longitudinal Survey of Youth 1979 child and young adult data are available in two data sets:

- *The NLSY79 child and young adult file* contains data and documentation, together with search and extraction software, for the NLSY79 1986–2002 child data collections and the 1994–2002 NLSY79 young adult surveys. This includes all child data elements described in chapter 4. Children and young adults can be linked to their NLSY79 mothers by case identification code. This data set is available on the cross-cohort CD or can be downloaded from the NLS Web site at www.bls.gov/nls/.
- *The NLSY79 young adult geocode CD* contains all data from the child and young adult public-use file, along with geocode data for the young adults from 1994 through 2002. Information on obtaining the young adult geocode CD can be found online at www.bls.gov/nls/geocodeapp.htm.

Original cohort data

The NLS Mature Women and Young Women data sets contain data and documentation, together with search and extraction software, for the 1967–2003 surveys of the Mature Women and the 1968–2003 surveys of the Young Women. These data sets are available on the cross-cohort CD or as separate files that can be downloaded from the www.bls.gov/nls Web site. Included are all of the data elements described in chapter 5, with the Mature Women pension plan data available as a separate file.

The NLS Older Men and Young Men data sets contain all of the information described in chapter 6 for the 1966–90 surveys of the Older Men and the 1966–81 surveys of the Young Men. These files are included on the cross-cohort CD or can be downloaded from the www.bls.gov/nls Web site.

NLS software

NLS Investigator search and extraction software is present on the cross-cohort CD and in all downloadable data files. The CD and download files also include installation and usage instructions. The Windows software features a user-friendly interface and allows the user to perform the following five functions:

1. **SEARCH** the NLS documentation files for categories of variables clustered by topical area. NLSY97, NLSY79, NLSY79 children, and Mature and Young Women searches can be based on (1) topical **area of interest** (Jobs, Income, Geo92, and so forth) or (2) **any word** or combination of words found within the variable titles (for example, “Activity [and] Survey [and] Week” or “Child [and] Age”). Due to differences in data structuring, search by area of interest is not currently available for the Older and Young Men; any word searches are available. Researchers can also choose specific variables by reference number, question number, question name, or look at all variables collected during a given survey year.
2. **VIEW** a screen image of each variable’s codebook information (including coding categories and frequencies for the full universe of respondents) or view data for a given respondent by identification code (for instance, case #260).
3. **SELECT** variables for inclusion in an extract specification file that will be used to generate the data and documentation files described below.
4. **EXTRACT** data for the full universe of respondents or for a sample restricted by race, sex, number of cases, or user-specified Boolean equation, and produce a raw ASCII data file (with SAS or SPSS control cards), a Dbase file, or a STATA dictionary file.
5. **DOCUMENT** an extracted file by producing a listing of all extracted variables, including each variable’s reference number, title, area of interest (if available), number of valid cases, and minimum-maximum-mean values, or a codebook file depicting the coding categories, full universe frequencies, verbatim question text and question number, reference number, and variable title.

More information about searching for variables of interest is provided in the user’s guides for the individual cohorts. A representation of typical Investigator extract screens appears in figure 8.2. Using the NLSY97 data, these example screens show a combined survey year and “any word” search for the respondent’s wage and salary income in 1998, as reported in round 3 (1999).

Codebook documentation

Each NLS cross-cohort CD and downloaded data set contains data, documentation in the form of codebook pages, and Investigator search and extraction software. The codebook is the principal element of the documentation system and, for each variable in the file, contains information intended to be complete and self-explanatory. In figure 8.2, the codebook information for the example wage and salary variable can be seen in the gray-shaded window labeled “NLS-Codebook.” For each variable, the codebook depicts a reference number, question title (and question text, in many cases), variable description, coding information, frequency distribution, and source. For the original cohorts and the NLSY79, derivations are provided for certain created variables. Codebooks for the original cohorts contain universe information for the early years and most recent surveys; separate flowcharts present universe information for the middle years. The codebook also lists the previous and next survey questions; for non-constructed variables, hyperlinks permit users to page through the codebook in the order that questions were asked in the interview.

The codebook contains references to pertinent attach-

Figure 8.2. Search and extraction screens for the NLS Windows software

The screenshot displays the CHRR NLS Investigator software interface. The main window title is "CHRR NLS Investigator: NLSY97 1997 - 1999 - SemiPrivate: [Tagset: wages99]". The interface is divided into several panes:

- Contents Pane:** Shows a tree view of the data structure, including "Documents" and "Indexes". The "Any Word in Context" index is selected.
- NLS Variable List - 26 variables:** A table listing variables with checkboxes for selection. The selected variables are:

N	Description	Reference Num...	Question Name	Area of Interest	Survey Year
<input type="checkbox"/>	INCOME IN WAGES, SALARY, TIPS FROM REGULAR OR ODD JOBS?	R36500.00	YINC-1500	INCOME	1999
<input checked="" type="checkbox"/>	TOTAL INCOME FROM WAGES AND SALARY IN PAST YEAR	R36502.00	YINC-1700	INCOME	1999
<input type="checkbox"/>	ESTIMATED INCOME FROM WAGES AND SALARY IN PAST YEAR	R36503.00	YINC-1800	INCOME	1999
<input checked="" type="checkbox"/>	SPOUSES TOTAL INCOME FROM WAGES AND SALARY PAST YEAR	R36511.00	YINC-2600	INCOME	1999
<input checked="" type="checkbox"/>	ESTIMATED SPOUSES TOTAL INCOME FROM WAGES AND SALARY PAST YEAR	R36512.00	YINC-2700	INCOME	1999
<input type="checkbox"/>	TOTAL INCOME FROM WAGES, BUSINESS, FARM IN PAST YEAR (BIO FATHER)	R36558.00	YINC-8800	PARENT RET...	1999
- NLS-Codebook/R3650200:** Provides detailed information for the selected variable R36502.00. It includes the question text: "During 1998, how much income did you receive from wages, salary, commissions, or tips from all jobs, before deductions for taxes or for anything else?" and a note about truncated values. Below this is a frequency distribution table:

Value	Frequency
43	0
833	1 to 499
479	500 to 999
422	1000 to 1499
174	1500 to 1999
303	2000 to 2499
82	2500 to 2999
187	3000 to 3499
42	3500 to 3999
148	4000 to 4499
31	4500 to 4999
261	5000 to 6999
423	7000 to 99999999: 7000+
3428	
- Search Results Table:** A table at the bottom left of the codebook pane showing search results for "Any Word in Context":

Item	Entries	Operator
INCOME	625	AND
WAGES	88	AND
1999	12146	AND

The status bar at the bottom indicates "26 variables selected."

The screenshot shows the "CHRRDBA: BINARY.DAT Extract" dialog box. It features several sections for configuring the extraction process:

- Buttons:** "Run Extract", "Stop Extract", "Write Codebook", and "Exit".
- Case Selection:**
 - Select 1 of every case(s).
 - First case to read:
 - Last case to read:
 - Button: "Edit Boolean Specification"
- Extract Data File Type:**
 - Formatted ASCII
 - Delimited ASCII
 - DBASE 3
 - Stata Dictionary
 - SAS Control Cards
 - SPSS Control Cards
- Optional Output Files:**
 - SAS Control Cards
 - SPSS Control Cards
 - Extract Codebook File
- Summary Statistics:**
 - Number of variables: 2
 - Number of cases: 8984
 - Output record length: 10
 - Output file size: 107808
 - Output file name: wages99.DAT

ments and appendixes that provide supplementary coding and variable creation procedures. For the NLSY79 child file, the codebook also includes source information on whether a given variable is created or is a direct pickup from one of the mother/child supplements.

In addition to viewing codebook pages electronically, users can generate and print codebook pages and variable lists using the Investigator software. This feature enables researchers to customize the documentation for their particular research needs and to select and print information for any or all pertinent variables.

NLS user’s guides and supplemental documentation

Extensive additional documentation is provided to assist researchers in using NLS data. Many items are available for download; some are distributed only in hard copy. When available, electronic versions of the questionnaire, codebook supplement, and user’s guide are included on the cross-cohort CD and with the downloaded data set. Documentation items can be ordered or downloaded from the product information page on the www.bls.gov/nls Web site. This page is the most accurate source for determining whether an item is

available in electronic or paper form.

Individuals ordering an NLS cross-cohort CD will automatically receive the following items:

1. Installation, usage, and maintenance instructions
2. Current paper editions of the user’s guide for the cohort of interest and the *NLS Handbook*
3. A free subscription to the quarterly NLS newsletter, *NLS News*, which is also available electronically at www.bls.gov/nls

Users downloading data from the www.bls.gov/nls Web site will receive the first two items listed electronically; the third is an option the user can choose while placing the order.

For detailed descriptions of the user’s guides and the NLS newsletter, see chapter 10 of this handbook. Information about viewing error notices on the Internet also is provided in chapter 10.

Available for each NLS data set is a set of supplemental documentation items, such as technical background reports, interviewer’s reference manuals, and various attachments and appendixes that provide coding information and variable creation procedures. Figure 8.3 lists the documentation items

Figure 8.3. NLS supplemental documentation items

Cohort	Documentation items described in this chapter (†) indicates that the item is updated after each survey round					
NLSY97	<p><i>Main/event history</i></p> <p>Technical Sampling Report</p> <p>Survey Instruments (†)</p> <p>Interviewer Reference Manuals (†)</p> <p>Codebook Supplement (†)</p>	<p><i>Geocode</i></p> <p>Geocode Codebook Supplement (†)</p>				
NLSY79	<p><i>Main</i></p> <p>Technical Sampling Report</p> <p>TSR Addendum (†)</p> <p>Household Screener and Interviewer’s Manual</p> <p>PAY–Attachment 106</p> <p>Addendum–AFQT Score Creation Procedure</p> <p>NLSY High School Transcript Survey</p> <p>Survey Instruments (†)</p> <p>Codebook Supplement (†)</p> <p>Interviewer Reference Manuals (†)</p> <p>Flowcharts (1979–92)</p>	<p><i>Work history</i></p> <p>Appendix 18 in main Codebook Supplement</p> <p><i>Geocode</i></p> <p>Geocode Codebook Supplement (†)</p>				
NLSY79 Child/ Young Adult	<p>Child and Young Adult Questionnaires (†)</p> <p>Child Assessment Tables (†)</p> <p>Child/Young Adult Data Users Guide (†)</p>	<p>NLSY79 Child Handbook</p> <p>NLSY79 Children 1992 Description and Evaluation</p> <p>Young Adult Attachments (†)</p>				
Original Cohorts	<p>Survey Instruments (†)</p> <p>Interviewer Reference Manuals/Field Representative Manuals (†)</p> <p>Flowcharts</p>	<p>Codebook Supplements (†)</p> <table style="width: 100%; border: none;"> <tr> <td style="width: 50%;">Mature Women</td> <td style="width: 50%;">Older Men</td> </tr> <tr> <td>Young Women</td> <td>Young Men</td> </tr> </table>	Mature Women	Older Men	Young Women	Young Men
Mature Women	Older Men					
Young Women	Young Men					

available for each cohort; detailed descriptions of these items follow in the same order. Items identified with a (†) are updated—either wholly or in part—for the most recent data release. Purchasers of any NLS data set are encouraged to access to all relevant documentation.

NLSY97 documentation

Supplemental documentation for the NLSY97 includes a number of different items. Documentation items that are new or updated for the latest release are identified with a (†). In addition, the NLSY97 user's guide is described in chapter 10.

NLSY97 main file and event history

Technical Sampling Report—Youth Survey. This technical manual, published by the National Opinion Research Center (NORC) at the University of Chicago, describes the procedures used to select the sample. The manual includes weights and standard errors for the initial survey year.

Survey instruments (†). This set of fielding materials serves to collect information from cohort respondents or other subjects of the survey during each interview. NLS User Services provides a new set of survey instruments, including a copy of the questionnaire and questionnaire supplements, for each survey year. The electronic codebook also incorporates questionnaire information.

Interviewer Reference Manuals (†). Included with each NLSY97 questionnaire is an interviewer's reference manual. This document reproduces computerized help screens that provide NORC interviewers with detailed question-by-question instructions for coding each survey year's questionnaire and other supplementary survey instruments.

Codebook Supplement (†). Reissued with each data release, the *Codebook Supplement* contains variable creation procedures and supplementary coding information not present in the electronic version of the codebook.

Attachment:

- 1. Census Industrial and Occupational Classification Codes.** This document lists the three-digit 1990 and 2002 census codes used to classify job and training information.

Appendixes:

- 1. Education and Training Variable Creation** (†). This document provides the programs for several created variables related to education and training. Education variables include enrollment status, type of school, date received diploma, highest grade completed, number of schools attended, and PIAT math score. Training vari-

ables indicate whether the respondent received a certificate or license as the result of a training program and the date on which the most recent certificate or license was received.

- 2. Employment Variable Creation** (†). This appendix provides programs for created employment variables, including hourly rate of pay, hourly monetary compensation, number of weeks worked, total tenure at job, and number of jobs held.
- 3. Family Background and Formation Variable Creation** (†). This appendix of created variable programs contains those dealing with family background, such as household size, marital status, fertility, and marriage and cohabitation history.
- 4. Geographic Variable Creation.** Several variables in the main data set provide information about the respondent's area of residence, permitting researchers to identify key characteristics of the area without needing the geocode CD. Included in this appendix is a summary of the four census geographic regions, an explanation of the metropolitan statistical area/central city status variable, and the definition for the rural vs. urban variable.
- 5. Income and Assets Variable Creation** (†). This document provides the creation procedures for income and assets variables. These include household net worth and gross household income, as well as receipt of public assistance.
- 6. Event History Creation and Documentation** (†). This appendix explains the structure of the event history variables and describes the creation process.
- 7. Continuous Month Scheme and Crosswalk** (†). This document explains the structure of the event history month-by-month and week-by-week status arrays and provides crosswalks from continuous month/week numbers to actual month and year dates.
- 8. Instrument Rosters** (†). In each survey round, a number of rosters are used to organize information during various parts of the interview. This appendix identifies these rosters and shows how they are used in different parts of the survey. It also lists the variable names, titles, and reference numbers for the various instrument rosters used in each round.
- 9. Family Process and Adolescent Outcome Measures.** This document, which is provided separately from the *Codebook Supplement*, summarizes the creation procedures for the various scales and indexes created by Child Trends, Inc. The appendix also presents the results of Child Trends' statistical analyses of the scales, indexes, and a number of related attitude and behavior variables.
- 10. CAT-ASVAB scores.** This appendix describes the administration of the CAT-ASVAB to NLSY97 respondents

during round 1. It also discusses the process used to create scores for each respondent who participated in the test.

- 11. Collection of the Transcript Data, Wave 1.** To help researchers understand the variables from respondent transcripts collected in round 3, this appendix describes the data collection instruments, data-entry and coding process, and resulting school- and respondent-specific variables.

NLSY97 geocode file

Geocode Codebook Supplement (†). This publication, which provides additional coding information not presented in the document codebook, contains attachments relating to the NLSY97 geocode CD and an introduction that discusses key aspects of the geocode CD. Brief descriptions of the contents of each document are provided below.

Attachments:

- 100. 1990 Census Bureau State and County Codes.** This attachment provides coding information for the State and county variables included on the NLSY97 geocode CD. These variables use the current Federal Information Processing Standards (FIPS) codes.
- 101. MSA Codes.** This document lists the Metropolitan Statistical Area (MSA) coding scheme used for NLSY97 geocode variables. The various sections also include Consolidated Metropolitan Statistical Area (CMSA) codes, New England Consolidated Metropolitan Area (NECMA) codes, and Primary Metropolitan Statistical Area (PMSA) codes.
- 102. IPEDS Data and College Identification Codes** (†). This attachment briefly describes the coding schemes used to identify the postsecondary educational institutions attended by respondents and lists codes assigned by survey personnel to institutions not represented in the Integrated Postsecondary Education Data System (IPEDS) coding schemes.

NLSY79 documentation

Documentation items for the NLSY79 files include various technical reports and round-specific documents that provide critical information on each data collection round. Those documentation items that are new or updated, either wholly or in part, for the latest release are identified with a (†).

NLSY79 main file

Survey design, sampling, and fielding procedures. The following documents provide technical information about the main NLSY79 and separately administered surveys, such as the 1980 high school survey and 1980–1983 transcript data collections.

Technical Sampling Report—Youth Survey. This technical manual, published by NORC in 1983, describes the selection procedures for the civilian and military youth samples. It includes weights and standard errors for the initial survey years. Updates to this document are released in an addendum.

Technical Sampling Report Addendum (†). This supplement to the 1983 *Technical Sampling Report* presents, in tabular form, the standard errors and design effects for each survey from 1982 through the latest round of the NLSY79.

Household Screener and Interviewer's Manual (1978). Used for the 1978 preliminary screening of households from which the two civilian youth samples were drawn, this manual contains instructions provided to NORC interviewers, question-by-question specifications for the household screener interview form, and a completed sample screener.

Profiles of American Youth—Attachment 106. This attachment provides general and technical information about the 1980 administration of the *ASVAB (Armed Services Vocational Aptitude Battery)* to NLSY79 respondents. Included are technical notes on the *ASVAB* scale scores, an annotated bibliography of U.S. Department of Defense publications, an example of the test score report, and various brochures disseminated to participating respondents.

Addendum—AFQT Score Creation Procedures. This addendum to Attachment 106 provides information on the creation of two Armed Forces Qualification Test scores, the AFQT80 and AFQT89, along with tables for converting subtest raw scores to standard scores. This set of variables was added to the NLSY79 data set beginning with the round 12 release.

NLSY High School Transcript Survey: Overview and Documentation. This document contains background information on the sample design, fieldwork, and types of variables collected during the three rounds of this special survey effort. It includes the transcript survey codebook, instructions for coding courses, course codes, and copies of the transcript coding form and school questionnaire. Finally, it presents references to other technical reports prepared by the sponsoring agency, the National Center for Research in Vocational Education.

Round-specific documents. Each data release includes either complete or partial updates to the following round-specific documents:

Survey instruments (†). The set of field materials used during each interview serves to collect information from cohort

respondents or other subjects of the survey. In years during which a paper-and-pencil interview was conducted, a new set of survey instruments was issued for each survey year and typically included a copy of the questionnaire, questionnaire supplements, the various versions of the household interview forms, the employer supplement, and a jobs calendar. In years during which computer-assisted personal interviewing was done, these items are all incorporated into one survey instrument, although a paper calendar is provided for reference.

Codebook Supplement (†). Reissued with each data release, the *Codebook Supplement* contains variable creation procedures and supplementary coding information. This information is contained in the following numbered attachments and appendixes:

Attachments:

3. **Industry and Occupation Codes.** This attachment compiles: (1) The three-digit 1970 census classifications used to code job and training information, as well as occupational aspiration information; (2) the three-digit 1980 census codes that have been used (in addition to the 1970 codes), beginning with the 1982 survey, to classify respondents' current or most recent job; (3) three-digit 2002 census codes used starting in 2004 to classify industry and occupation for all jobs; and (4) the 1977 military occupational specialty codes used to classify responses to the 1979–85 questions on military jobs and military occupations.
4. **Fields of Study in College.** This document provides the coding classifications for the variables regarding major fields of study and subspecialties at colleges attended.
5. **Index of Labor Unions and Employee Associations.** This index provides codes for the 1979 questions regarding the name of the union/employee association at jobs #1 through #5 (for example, R00937.–R00941.).
6. **Other Kinds of Training Codes.** Listed in this attachment are the various categories of occupational training used to code the 1979 survey question on types of other training programs in which a respondent was enrolled for at least 1 month (R01348., R01353., R01358., R01363.).
7. **Other Certificates Codes.** This list defines codes for the various types of certifications (practical nurse, welding, insurance, chef, and so forth) that a respondent had ever received as of the 1979 interview (R01376., R01377., R01378., R01379.).
8. **Health Codes.** This modified version of the International Classification of Diseases (ICD-9) codes (*International Classification of Diseases*, Volumes 1 and 2, Geneva, WHO, 1977–78) was used to classify types of health problems limiting the amount or kind of work that a respondent could do (1979–82 surveys) or occurring as a result of a work-related injury or illness (1988–90 and 1992–2000 surveys). These codes also were used in 1998–2002 to classify major health problems of biological parents for respondents 40 years or older. Also included is a list of numeric codes identifying the parts of the body affected by health problems or work-related injury or illness.
100. **Geographic Regions.** This attachment lists the States making up each of the four regions used in 'South/Non-South Place of Birth,' 'South/Non-South Place of Residence at Age 14,' and such yearly variables as 'Region of Residence.'
101. **Country Codes.** Provided in this document are the foreign country codes for respondents' country of residence, country of parents' birthplace, and country of citizenship at the time of immigration.
102. **State FIPS Codes.** The Federal Information Processing Standards are used to code respondents' State of birth and State of residence.
103. **Religion Codes.** This list contains the various denominational categories used to code the religion questions in 1979 (R00103.10 and R00104.10) and 1982 (R06558., R06583., R06586., R06613., and R06616.).

Appendixes:

1. **Employment Status Recode (ESR) Variable Creation 1979–1998** (†). This appendix provides the adapted PL/I version of the U.S. Census Bureau's FORTRAN program used to create the measure of main labor force activity during the survey week from 1979–1993. Beginning in 1994, the SPSS formula for creating the variable is provided; this formula corresponds closely to the revised formula provided by the U.S. Census Bureau for the Current Population Survey (CPS).
2. **Total Net Family Income Variable Creation 1979–2002** (†). Contained in this appendix is the PL/I code for 1979–1996 and the SPSS code for 1998–2002 used to create the *KEY* income variable for each survey year, as well as the poverty level and poverty status variables.
3. **Job Satisfaction Measures 1979–1982 and 1988.** This document provides background information and yearly reference numbers for both the scale items and global satisfaction measures of the modified Quality of Employment Survey scale administered in the 1979–1982 and 1988 surveys. It includes additional references and a methodology for constructing the full scale.
4. **Job Characteristic Index 1979 and 1982.** This appendix gives background information, reference numbers, questionnaire locations, and additional references for the job complexity questions asked in these 2 survey

- years.
5. **Supplemental Fertility File Variables** (†). This discussion provides a brief overview of the constructed variables in the 1979–92 and 1994–2002 Fertility and Relationship History/Created area of interest on the main NLSY79 data file, background information on the 1982 and 1994 data quality checks, and information on the availability of additional reports assessing the NLSY79 fertility data.
 6. **SMSA Urban-Rural Creation** (†). Contained in this appendix are the decision rules used to create the ‘Current Residence in SMSA’ variables and the ‘Is R’s Current Residence Urban/Rural?’ variable series.
 7. **Unemployment Rate** (†). This discussion explains how the variable ‘Unemployment Rate of Labor Market of Current Residence’ is created.
 8. **Highest Grade Completed and Enrollment Status Variable Creation** (†). This document contains the code used to create the *KEY* 1990–2002 variables ‘Highest Grade Completed as of May 1 Survey Year’ and ‘Enrollment Status as of May 1 Survey Year.’ It includes a discussion of the revised versions of the 1979–2002 ‘Highest Grade Completed’ variables, which were released for the first time with the 1994 data.
 9. **Linking Employers through Survey Years.** This appendix identifies the procedures and variables necessary for linking employers reported across contiguous interview years.
 11. **NLSY79 Round 12 (1990) Survey Administration Methods.** This appendix briefly describes the computer-assisted personal interviewing (CAPI) and paper-administered personal interviewing (PAPI) techniques employed during the 1990 NLSY79 interviews.
 12. **Most Important Job Learning Activities (1993–1994).** Provided in this document are value labels for the eight 1993/1994 variables that identify method(s) used by respondents in learning to perform job duties associated with their current or most recent job.
 13. **Introduction to the 1993 through 2002 CAPI Questionnaires and Codebooks** (†). This discussion introduces readers to the new documentation items and terms used in the CAPI survey instruments beginning in 1993 and alerts the users of these data to important changes in the way some variables are coded, documented, and generated.
 14. **1993–2002 Instrument Rosters** (†). This consolidated list provides subject content and matching reference numbers for various types of roster items occurring in the CAPI surveys.
 15. **Reciprocity Event Histories** (†). This appendix describes the logic of the new CAPI event history format used to collect information on unemployment, AFDC/TANF, Food Stamps, and other welfare reciprocity. It also describes the process whereby the RECIPIENCY event histories (month-by-month and yearly) are created.
 16. **1994 Recall Experiment.** This discussion explains how the 1994 recall experiment was designed to examine the effects on the accuracy and consistency of respondents’ recall when the frequency of survey administration was changed from annual to biennial. A subsample of NLSY79 respondents who were interviewed in both 1992 and 1993 answered questions in 1994 referring to the period since their 1992 interview. These data were then compared with respondents’ answers to the 1993 survey for possible discrepancies.
 17. **Interviewer Characteristics Data.** For each round, this appendix describes the data available regarding the characteristics of NLSY79 interviewers; this information is based on NORC’s interviewer personnel files. These data permit researchers to link interviewers with the respondents they interviewed.
 18. **Overview of the NLSY79 Work History Data.** This appendix introduces users to the work history data. It describes the creation of the data set, the programs used to create it, and the input variables.
 19. **SF-12 Summary Scores.** This appendix explains the administration of the SF-12 health scale, a brief inventory of self-reported mental and physical health. This was administered to respondents who had turned 40 since their last interview as part of the age 40+ health module.
- Interviewer Reference Manuals* (†). Accompanying each NLSY79 questionnaire is an interviewer’s reference manual, called the *Question-by-Question Specifications (Q by Q)*. This document provides NORC interviewers with background information on the NLSY79 and, where necessary, detailed question-by-question instructions for coding each survey year’s questionnaire, employer supplement, household interview forms, and other supplementary survey instruments. Beginning with the 1993 interviews, copies of the interviewer CAPI help screens are incorporated as part of each *Q by Q*. All users must have access to a copy of the latest *Q by Q*.
- Flowcharts.* These schematic diagrams, available for the 1979–92 surveys, depict the often complex skip patterns in the main questionnaire and questionnaire supplements.
- NLSY79 work history documentation*
Appendix 18, “Overview of the NLSY79 Work History Data” (found in the main NLSY79 codebook supplement), introduces users to the work history data. It describes the creation of the data set, the programs used to create it, and the input variables.

NLSY79 geocode file

Geocode Codebook Supplement (†). This publication, which provides additional coding information not presented in the electronic codebook, contains attachments and appendixes relating to the NLSY79 geocode CD and an introduction that discusses key aspects of the geocode CD. Brief descriptions of the contents of each document are provided below. Items identified with a (†) are updated for the latest release of data.

Attachments:

- 100. Geographic Regions.** This document lists the States making up each of the four regions used in coding such variables as ‘Region of Residence,’ ‘South/Non-South Place of Birth,’ and ‘South/Non-South Place of Residence at Age 14.’
- 101. Country Codes** (†). Provided in this document are the foreign country codes for respondents’ country of residence, country of parents’ birthplace, and country of citizenship at the time of immigration.
- 102. State FIPS Codes.** The Federal Information Processing Standards are used to code respondents’ State of birth and State and county of residence.
- 104. SMSA Codes** (†). This attachment contains coding information used to classify the respondent’s SMSA, MSA, CMSA, or PMSA of residence at each interview date.
- 105. Addendum to FICE Codes** (†). This document provides supplementary identification codes for colleges and universities not listed in the *Education Directory: Colleges and Universities* (1981–82 and 1982–83 supplement) published by the National Center for Education Statistics. It includes information on the revised Federal Interagency Committee on Education (FICE) code variables first present on the 1994 release.

Appendixes:

- 7. Unemployment Rate** (†). This discussion explains how the continuous and collapsed versions of the variable ‘Unemployment Rate for Labor Market of Current Residence’ were created.
- 10. Geocode Documentation** (†). This appendix provides background information on the creation of the original 1979–82 geocode tape and later updates on the modification of those data for the 1979–2002 release.

NLSY79 child and young adult documentation

Documentation for the child file includes survey instruments (often called “supplements”), a child and young adult data users guide, a child handbook covering assessment procedures and data collected from 1986–90, and summary reports evaluating data collections since 1990. Those items identified with a (†) are updated, either wholly or in part, for the

latest data release. Researchers using the NLSY79 child file who make extensive use of data derived from the mother’s record should consider ordering the relevant NLSY79 main file documentation items as well as the child materials.

Other summary reports also are available to researchers working with the child data; chapter 10 of this document provides descriptions and ordering information.

Child Assessment Instruments (†). Depending on the year, the child survey includes a *Child Supplement*, a *Child Self-Administered Supplement* for children 10 years of age and older, and a *Mother Supplement*. The mode of administration—paper or CAPI—also varies depending on the year.

Young Adult Questionnaire (†). The young adults (NLSY79 children age 15 and older) are administered the CAPI *Young Adult Questionnaire*. Through 1998, they were administered the paper *Young Adult Survey—Self-Report Booklet* as well. The *Young Adult Self-Report* was a paper booklet until 2000, when the young adult self-report questions were incorporated into the CAPI questionnaire.

Child Assessment Tables (†). These year-specific reports contain descriptive tables based on the NLSY79 child assessment data. They provide detailed distributions of raw and normed scores by age and race or ethnicity for overall scales, as well as for any relevant subscales. The tables are available for each child assessment year starting with 1994.

Child and Young Adult Data Users Guide (†). This guide is updated for each data release. It provides information on the contents of the data, characteristics of the child and mother samples, sampling constraint issues of which users should be aware, and a discussion of each child assessment administered during the interviews. Users are encouraged to access the 2000 guide as well as the 2002 data guide if they plan to use data from more than one survey round because the current guide is not comprehensive. With the introduction of the *Young Adult Survey* in 1994, the content of the *Young Adult Questionnaire* is discussed in detail. The guide also outlines potential research applications that exploit the longitudinal and cross-cohort potential of the child and young adult data.

NLSY79 Child Handbook (1986–90). This document summarizes the administration of child assessments during the 1986–90 period.

NLSY79 Children 1992 Description and Evaluation. This publication describes the 1992 child assessment data, including completion rates for assessments by age, race, and ethnicity. It offers in-depth discussions of outcomes over time and evaluates the validity of the assessments.

NLSY79 young adult attachments (†). These numbered attachments provide supplementary coding information not contained in the codebook, as well as the electronic question-by-question specifications available to the interviewers. Through 2000, these attachments were included in the printed questionnaire rather than as a separate booklet. Beginning in 2002, they are provided as HTML documents that are downloadable with the data.

Attachments:

3. **1970 Census Industry/Occupation Codes.** This attachment contains the three-digit 1970 census classifications used to code job and training information and *Employer Supplements* (U.S. Census Bureau, *1970 Census of Population Alphabetical Index of Industries and Occupations*, U.S. Government Printing Office, Washington, DC, 1971).
4. **1990 Census Industry/Occupation Codes.** This document lists the three-digit 1990 census classifications used for double coding of occupation and industry for the CPS job (U.S. Census Bureau, *1990 Census of Population Alphabetical Index of Industries and Occupations*, U.S. Government Printing Office, Washington, DC, 1991).
5. **Electronic Question-by-Question Specifications.** This attachment lists the help screens available to the interviewer during the administration of the CAPI questionnaire and indicates the question numbers that were linked to these screens.
6. **2000 Census Industry/Occupation Codes.** This document lists the three-digit 2000 Census classifications used for the coding of occupation and industry for all jobs in the 2002 survey data (U.S. Census Bureau, www.census.gov/hhes/www/ioindex/ioindex.html).

NLS original cohort documentation

Documentation items for each original cohort data file are specific to the cohort. Relevant documentation includes a set of questionnaires, a codebook supplement, interviewer reference manuals, and flowcharts. Those items identified with a (†) are updated, either wholly or in part, for the most recent release of data for each cohort.

Sets of survey instruments (†). A unique set of survey instruments serves to collect information from respondents during each interview. Each cohort has a separate set of questionnaires; new questionnaires are issued for each survey year. A set of survey instruments includes not only copies of each survey's questionnaire, but also copies of the household record cards used before 1995 to record information on the composition of each respondent's household.

Interviewer Reference Manuals/Field Representative's

Manuals (†). Accompanying each questionnaire is a field reference manual that provides Census Bureau field representatives with background information on the NLS, respondent locating instructions, and detailed question-by-question instructions for coding and otherwise completing the questionnaire and household record cards.

Flowcharts. Universe information and skip patterns for 1977–93 questionnaires are provided within separate flowcharts, available for purchase by single year or by set. The cohort codebooks contain comparable information for earlier years. For CAPI administrations of the surveys of women beginning in 1995, this type of information is incorporated in the questionnaire.

Codebook Supplements (†). This set of numbered attachments and appendixes contains cohort-specific supplementary coding and variable creation information not present in the codebook. A listing of the contents of each codebook supplement follows below. Contents of the Mature Women supplement appear first, followed by those for the Young Women, Older Men, and Young Men.

Mature Women codebook supplement

Attachments:

2. **Census of Population Industry and Occupational Classification Codes.** This document provides the occupation-industry coding assignments made by Census Bureau personnel from the verbal descriptions obtained in the interviews. Codes from 1960, 1980, 1990, and 2000 have been used with this cohort; table 8.1 summarizes the survey years and jobs for which each scheme is available. This attachment also contains a copy of the Duncan Socioeconomic Index, an ordinal prestige scale assigning a rank of 0 through 97 to each of the three-digit 1960 census occupations.
4. **Bose Index.** This list provides a mean occupational prestige score for each of the three-digit 1960 occupation codes for the Mature and Young Women cohorts.
5. **Employment Status Recodes.** This attachment describes the methodology used by the Census Bureau to calculate each respondent's employment status from the CPS questions asked in each NLS interview. It provides (1) definitions of 'working,' 'with a job but not at work,' 'unemployed,' and 'not in the labor force'; (2) the decision rules used to assign or recode respondents to a particular labor force status; and (3) Census Bureau methodology for dealing with exceptions to the rules.

Appendixes:

3. State Names and State Codes by Census Division Listing

4. Derivations for R02847. (Reason left Current Job 1967)
5. Source for Occupational Atypicality Codes
6. Derivations for R02872.50 (Occupational Training 1967–1972)
7. Derivations for R00792.50 (Training Prior to 1972)
8. Derivations for R00744.05 (Number of Years Worked Before 1967 in Occupation)
9. Derivations for 1977 *KEY* Variables
10. New Geographical and Environmental Variables 1967–1977
11. Derivations for 1971 *KEY* Variables
12. Derivations for 1972 *KEY* Variables
18. Union Categories—Copy of Coding Instructions for Name of Union or Employee Association
19. Derivations for 1982 *KEY* Variables
20. Derivations for 1987 *KEY* Variables
21. Derivations for 1989 *KEY* Variables
22. Derivations for 1992 *KEY* Variables
23. Geometric Progression Coding (†)
24. Pension Plan Data Documentation (†)
36. Summary of the Major Differences between the 1995 and Earlier Surveys
37. Summary of 1995 Data Cleaning Issues
38. Derivations for 1995 *KEY* and Other Created Variables
39. Summary of 1997 Data Cleaning Issues
40. Derivations for 1997 *KEY* and Other Created Variables
41. Derivations for 1999 *KEY* and Other Created Variables
42. Derivations for 2001 *KEY* and Other Created Variables
43. Derivations for 2003 *KEY* and Other Created Variables

Young Women codebook supplement

Attachments:

2. **Census of Population Industry and Occupational Classification Codes.** This document provides the occupation-industry coding assignments made by Census Bureau personnel from the verbal descriptions obtained in the interviews. Codes from 1960, 1980, 1990, and 2000 have been used with this cohort; table 8.1 summarizes the survey years and jobs for which each scheme is available. This attachment also contains a copy of the

Duncan Socioeconomic Index, an ordinal prestige scale assigning a rank of 0 through 97 to each of the three-digit 1960 census occupations.

4. **Bose Index.** This list provides a mean occupational prestige score for each of the three-digit 1960 occupation codes for the Young Women and Mature Women cohorts.
5. **Employment Status Recodes.** This attachment describes the methodology used by the Census Bureau to calculate each respondent's employment status from the CPS questions asked in each NLS interview. It provides (1) definitions of 'working,' 'with a job but not at work,' 'unemployed,' and 'not in the labor force'; (2) the decision rules used to assign or recode respondents to a particular labor force status; and (3) Census Bureau methodology for dealing with exceptions to the rules.

Appendixes:

1. Fields of Study in College—Instructions for the Coding Scheme
2. State Names and State Codes by Census Division Listing
4. Listing of Median Education for Different Occupations
5. Source for Occupational Atypicality Scores
6. Supplemental Edit Specifications for *KEY* Variables: R03297., R03292., R03294., R03293., R03295.
7. Listing of Correction to Employment Status Recode for 1968 and 1969
9. Determinants of Early Labor Market Success: App. A
10. Determinants of Early Labor Market Success: App. B
11. Determinants of Early Labor Market Success: App. C
12. Determinants of Early Labor Market Success: Method for Variable Construction
18. Union Categories—Copy of Coding Instructions for Name of Union or Employee Association
20. Derivations for R05007., R05012. (Marital Status Patterns)
21. Rules for Revising Variables Rep (Month/Yr Since Left School)
22. GED/SVP/Job-Level/Job Family Values

Table 8.1. Mature and young women industry and occupation coding systems by survey year

Coding system	Mature women	Young women
1960 codes	1967–92	1968–93
1980 codes—current/last job only	1984, 1986	1983–87
1980 codes—current/last job and dual job only	1987–92	1988, 1991
1980 codes—all jobs	1995–99	1993–99
1990 codes—current/last job and dual job only	1992	—
1990 codes—all jobs	1995–99	1993–99
2000 codes—all jobs	1995–2003	1995–2003

23. Derivations for R05031.–R05047. (Occupation and Other Job Information before Birth of Child)
24. Derivations for R05049.–R05060. (Occupation and Other Job Information after Birth of Child)
25. New Geographic and Environmental Variables for 1968–1978
26. Derivations for 1978 *KEY* Variables
27. Source for the Job Characteristics Index
28. Source for the Job Satisfaction Measures
29. Reason for Preference in Union Certification Election (Item 10e, 1982, R07627.)
30. Derivations for the 1983 *KEY* Variables
31. Listing of Changes in 1983 Survey Made after Questionnaire Printed
32. Derivations for the 1988 *KEY* Variables
33. Derivations for the 1991 *KEY* Variables
34. Derivations for the 1993 *KEY* Variables
35. Geometric Progression Coding
36. Summary of the Major Differences between the 1995 and Earlier Surveys
37. Summary of 1995 Data Cleaning Issues
38. Derivations for 1995 *KEY* and Other Created Variables
39. Summary of 1997 Data Cleaning Issues
40. Derivations for 1997 *KEY* and Other Created Variables
41. Derivations for 1999 *KEY* and Other Created Variables
42. Derivations for 2001 *KEY* and Other Created Variables
43. Derivations for 2003 *KEY* and Other Created Variables

Older Men codebook supplement

Attachments:

2. **Census of Population Industry and Occupational Classification Codes.** This document provides the occupation-industry coding assignments made by Census Bureau personnel from the verbal descriptions obtained in the interviews. The 1960 census classifications have been used to code all occupation and industry variables for all survey years. In addition, the Census Bureau double coded the current or last job held by the respondent using the 1960 and 1980 classifications in 1983. This attachment also contains a copy of the Duncan Socioeconomic Index, an ordinal prestige scale assigning a rank of 0 through 97 to each of the three-digit 1960 census occupations.
5. **Employment Status Recodes.** This attachment describes the methodology used by the Census Bureau to calculate each respondent's employment status from the CPS questions asked in each NLS interview. It provides (1) definitions of 'working,' 'with a job but not at work,' 'unemployed,' and 'not in the labor force'; (2) the decision rules used to assign or recode respondents to a particular labor force status; and (3) Census Bureau methodology for dealing with exceptions to the rules.

Appendixes:

1. Table Used for Determining Cutoff Points for Poverty Status Variables
2. Additional Instructions to Interviewers for the 1969 Survey
3. State Names and State Codes by Census Division Listing
4. Explanation for Census Processing of Weeks on Layoff (R00737. in 1967 and R01248. in 1969)
5. New Geographic and Environmental Variables 1966–1976
18. Union Categories—Copy of Coding Instructions for Name of Union or Employee Association
19. Derivations for 1981 *KEY* Variables
27. Procedures for Calculating the 1990 Weights (R06013.10, R06013.20, R07113.10, R07113.20, R07113.30)
28. Derivations for 1990 *KEY* Variables
29. Summary of Major Differences Between the 1990 and Earlier Surveys
30. Explanation for the 1990 Created Social Security Variables (R06743.10–R06743.30)
31. Derivations for Collapsed Occupation and Industry Variables (R07092.–R07098.)
32. Procedures for Coding Cause of Death—Supplemental Death Certificates

Young Men codebook supplement

Attachments:

2. **Census of Population Industry and Occupational Classification Codes.** This document provides the occupation-industry coding assignments made by Census Bureau personnel from the verbal descriptions obtained in the interviews. The 1960 census classifications have been used to code all occupation and industry variables for all survey years; only the 1960 codes were used with this cohort. This attachment also contains a copy of the Duncan Socioeconomic Index, an ordinal prestige scale assigning a rank of 0 through 97 to each of the three-digit 1960 census occupations.
5. **Employment Status Recodes.** This attachment describes the methodology used by the Census Bureau to calculate each respondent's employment status from the CPS questions asked in each NLS interview. It provides (1) definitions of 'working,' 'with a job but not at work,' 'unemployed,' and 'not in the labor force'; (2) the decision rules used to assign or recode respondents to a particular labor force status; and (3) Census Bureau methodology for dealing with exceptions to the rules.

Appendixes:

1. Fields of Study in College—Instructions for the Coding Scheme
2. State Names and State Codes by Census Division Listing

3. Codes for Enlisted Men and Officers in the Military
4. Listing of Median Education for Different Occupations
5. Listing of Corrections to Employment Status Recode for 1966, 1967, 1968, and 1969
6. Listing of the Method Used to Create Respondent's Father and Mother 1966 Variables
7. Derivatives for Creation of Dropped Variable Used to Create R02958.
8. Derivations Used to Create Enrollment Pattern 1966–1969
9. Determinants of Early Labor Market Success: App. A
10. Determinants of Early Labor Market Success: App. B
11. Determinants of Early Labor Market Success: App. C
12. Determinants of Early Labor Market Success: Method for Variable Construction
13. Dual Labor Market Theory: Appendix A (Primary and Secondary Jobs)
14. Edits for R02997.
15. Decision Criteria for Construction of Variables for Date Last Attended School
16. Listing of Corrections Made for 21 Variables (R02315.–R02317., R02320.–R02322., R02611.–R02622.)
17. Work Experience Variables: Time Reference for the 1970 Work History Section
18. Union Categories—Coding Instructions for Name of Union or Employee Association
19. GED/SVP/Job-Level/Job-Family and 3 Digit Occupation Code
20. Derivation for R03836. and Number of Weeks *KEY* Variables Between 1975 and 1976
21. New Geographic and Environmental Variables 1966–1976
22. Source for the Job Characteristics Index
23. Source for the Job Satisfaction Measures
24. Derivations for R03035., R03829., R03833.
25. Reason for Preference in Union Certification Elections
26. Derivations for R08114.–R08118.