



## Cohort Profile

# Cohort Profile: The National Longitudinal Survey of Youth 1979 (NLSY79)

Donna S Rothstein,<sup>1</sup> Deborah Carr<sup>2\*</sup> and Elizabeth Cooksey<sup>3</sup>

<sup>1</sup>U.S. Bureau of Labor Statistics, Washington, DC, USA, <sup>2</sup>Department of Sociology, Boston University, Boston, MA, USA and <sup>3</sup>CHRR at the Ohio State University, Columbus, OH, USA

\*Corresponding author. Department of Sociology, Boston University, 100 Cummington Mall, Boston, MA 02215, USA.  
E-mail: carrds@bu.edu

Editorial decision 24 May 2018; Accepted 2 June 2018

### Why was the cohort set up?

The National Longitudinal Survey of Youth 1979 (NLSY79) is a cohort of men and women born in the USA between 1957 and 1964, who were aged 14 to 22 when first interviewed in 1979. The survey was established to study the transition from school to work and careers among adolescents and young adults in the late 1970s. An additional goal of the survey was to help policy makers and researchers evaluate the expanded employment and training programmes for youths legislated by the 1977 amendments to the Comprehensive Employment and Training Act (CETA).

The NLSY79 is part of a larger family of longitudinal surveys sponsored by the U.S. Department of Labor (DOL), and was designed to compare the 1979 cohort's educational and labour market experiences with those of earlier cohorts who participated in the National Longitudinal Survey (NLS) programme's predecessor surveys. The NLS programme began in the 1960s with four cohorts of US adults: Older Men [born (b.) 1906–21], Mature Women (b. 1922–37), Young Men (b. 1941–52), and Young Women (b. 1943–54). In 1997, the NLS programme expanded with the establishment of the NLSY97, a cohort of youth born between 1980 and 1984, who were interviewed for the first time in 1997.

The NLSY79 has been expanded to facilitate the study of maternal characteristics and their impact on child development, health and well-being. In 1986, a companion

study that comprises the biological children of female participants in the NLSY79 began. This companion study was initially called the NLSY79 Child survey and included children aged 14 and younger. In 1994, the NLSY79 Young Adult survey was added to assess the experiences of these children as they turned age 15.

### Who is in the cohort?

The NLSY79 consists of three independent probability samples that total 12 686 respondents, representative of men and women who were born between 1957 and 1964 and were residing in the USA in 1979. The first is a cross-sectional sample of 6111 civilians. The second includes 5295 civilian men and women in three oversamples: (i) Black, non-Hispanic; (ii) Hispanic; and (iii) economically disadvantaged non-Black non-Hispanic. A third sample of military youth, funded by the U.S. Department of Defense (DOD), comprises 1280 men and women who were born between 1957 and 1961 and were serving in the military as of 30 September 1978. The original sampling was conducted by dwelling unit, and all members of each randomly selected unit who met the age criteria were selected to be part of the sample. Consequently, about half of all NLSY79 participants have at least one sibling in the sample, allowing explorations of within-family similarities and differences. Additionally, the biological children of all female respondents make up the companion NLSY79

Child and Young Adult surveys. These children were born between 1970 and 2011 and enable unique cross-generation analyses, although they do not constitute a cohort. More than 11 500 children have been born to female respondents of the NLSY79.<sup>1</sup>

### How often have they been followed?

NLSY79 respondents were first interviewed in 1979 and have been interviewed 27 times, annually until 1994 and biennially since then. The data span nearly 40 years and respondents will be ages 54 to 61 in round 28, to be undertaken in 2018. The military sample was dropped after the 1984 interview, and the economically disadvantaged non-Black, non-Hispanic oversample was dropped after the 1990 interview, leaving 9964 respondents eligible for interviews. Interviews were administered in person from 1979 through 1986, via telephone in 1987 and in person again from 1988 through 2000. Since 2002, most interviews have been conducted via telephone.

Response rates in the NLSY79 are high, as shown in Table 1.<sup>2</sup> Interviews were completed with 89% of those who were selected for base year (1979) interviews.<sup>3</sup> Deceased respondents are excluded from the denominator for the response rates in this table. Response rates remained above 90% through the first 16 rounds (up through the 1994 interview), and remained above 80% through the 2010 interview, or the first 24 rounds of the survey. Longitudinal surveys such as the Panel Study of Income Dynamics (PSID) and Survey of Income and Program Participation (SIPP) 2008 panel have considerably lower response rates (56% after 20 years in 1988 and under 55% after 5 years in 2013, respectively).<sup>4</sup> Attrition in the first 13 rounds of the NLSY79 (up to 1991) does not have a substantial influence on estimates from models of employment and wages, and researchers find no evidence that attrition in the first 26 rounds of data (through 2014) biases the relationship between labour market outcomes at age 30 and educational attainment or other characteristics.<sup>5,6</sup>

### What has been measured?

Box 1 provides a partial snapshot of the extensive and diverse topics covered in the survey. The NLSY79 began when respondents were in their teens and early 20s and has followed them through their late 50s and early 60s, so the survey content has evolved accordingly. The main focus has been labour force participation and experiences, with extensive and repeated measures of employment status and characteristics of each job held such as wages, hours and benefits, as well as detailed information on education, training, marriage, fertility, income, assets,

**Table 1.** Response rates in the NLSY79

Year	<i>n</i> response	Total deceased	Response rate (%)
1979	12 686	0	–
1980	12 141	9	95.8
1981	12 195	29	96.3
1982	12 123	44	95.9
1983	12 221	57	96.8
1984	12 069	67	95.6
1985	10 894 <sup>a</sup>	79	94.5
1986	10 655	95	92.6
1987	10 485	110	91.2
1988	10 465	127	91.2
1989	10 605	141	92.5
1990	10 436	152	91.1
1991	9018 <sup>b</sup>	144	91.9
1992	9015	156	91.9
1993	9011	177	92.1
1994	8891	204	91.1
1996	8636	243	88.8
1998	8399	275	86.7
2000	8033	313	83.2
2002	7724	346	80.3
2004	7661	399	80.1
2006	7654	456	80.5
2008	7757	503	82.0
2010	7565	573	80.6
2012	7301	689	78.7
2014	7071	790	77.1

Response rate is defined as the percentage of base-year respondents remaining eligible and not known to be deceased, who were interviewed in a given survey year.

<sup>a</sup>A total of 201 military respondents were retained from the original sample of 1280; 186 of the 201 participated in the 1985 interview. The total number of NLSY79 civilian and military respondents eligible for interview (including deceased respondents) beginning in 1985 was 11 607.

<sup>b</sup>The 1643 economically disadvantaged non Black/non-Hispanic male and female members of the supplemental subsample were not eligible for interview as of the 1991 survey year. The total number of NLSY79 civilian and military respondents eligible for interview (including deceased respondents) beginning in 1991 was 9964.

government programme participation and both spouse and child characteristics. A unique aspect of the NLSY79 is the availability of cognitive test scores obtained both in adolescence and at age 48. Respondents took the Armed Services Vocational Aptitude Battery (ASVAB) in 1980, and the maths and verbal components form what is known as the Armed Forces Qualification Test (AFQT) score. Age 48 cognitive measures include self-rated memory, word recall and backward-counting exercises. The NLSY79 also has tracked place of residence in every round; these data are available at the county level on a dataset accessible to researchers at their institutions, through arrangement with the Bureau of Labor Statistics (BLS), or at the census tract level on a dataset accessible at the BLS National Office or in a Federal Statistical Research Data Center. These data

**Box 1. Summary of NLSY79 content****Jobs**

Wage, hours, occupation, industry, tenure, benefits, promotions, job characteristics, job responsibilities, shift, class of worker, union status, benefits, reason left job, start and stop dates, dates of non-work during job, business ownership, military, job search.

**Labour market status**

Weeks working, unemployed and out of the labour force.

**Education**

Enrolment status, high school information, highest grade completed, highest degree, college type, college major, college loans and financial aid, high school transcripts.

**Training**

Type of training, whether completed, vocational degree or certificate, hours, dates of training, reason for training, financing, informal job training.

**Cognition**

ASVAB and AFQT scores (1980), age 48 cognitive measures including memory self-assessment, word recall and backward counting by sevens.

**Marriage and cohabitation**

Marital status and dates of changes in status, cohabitation information (1990 and on), employment information and other characteristics of spouse/partner.

**Fertility and children**

Pregnancy history, pre- and postnatal behaviours, children's dates of birth and sex. Extensive information about children of female respondents, including child cognitive and non-cognitive assessments through the Children of the NLSY79 survey.

**Health**

Limitations on work, self-rated health, height, weight, insurance coverage, health behaviours, physical activity, depressive symptoms, health conditions, extensive health modules at ages 40, 50 and 60.

**Personality and non-cognitive skills**

Internal-external locus of control scale, mastery scale, self-esteem scale, 10-item personality inventory, risk aversion measures.

**Programme participation**

Timing and amounts from unemployment insurance, public assistance, food stamps and more (self and spouse/partner).

**Income and assets/debt**

Family income, earnings for self and spouse/partner, child support, pensions, asset holdings, debt, bankruptcy, home foreclosure, financial literacy, wills.

**Smoking, drugs and alcohol use**

Cigarette use, extent of drug use, alcohol consumption.

**Childhood and family background**

Race/ethnicity, birthplace, religious affiliation, language in home, household resources at age 14, parental education, birthplace and other background.

**Residence history**

Residential information available every round, migration history, urban/rural, region.

**Retirement planning and expectations**

Respondent definition of retirement for self, expectations about age will stop working, retirement preparation activities, expectations about longevity.

allow researchers to append contextual variables defined at a census geography level.

The health measures have evolved over time too, reflecting age-related changes in health care needs and symptoms. General health and health limitations with respect to employment are assessed in every round. Mental health, health insurance coverage, health-related behaviours and medical treatments are assessed periodically, as are alcohol and drug use, smoking and body weight. When respondents turn ages 40, 50 and 60, the health modules are expanded with age-appropriate measures such as chronic health symptoms and conditions, functional limitations, access to care and service use. These expanded modules include widely used and validated measures such as the SF-12 scale and the Center for Epidemiological Studies Depression Scale (CESD), as well as self-reported assessments of major illnesses (e.g. asthma, cancer, diabetes, hypertension), cancer site, cognitive functioning, sleep problems, medication use, history of head injury, retrospective reports of early life illness or injury and major health problems of biological parents.

**What has it found? Key findings and publications**

The NLSY79 is a major resource for researchers who study employment, education, cognitive and non-cognitive skills, family and health, among other subjects.<sup>7</sup> As of December 2017, more than 2400 journal articles and 740 PhD

dissertations have used the data. A searchable bibliography is available at [<https://www.nlsinfo.org/bibliography-start>]. The NLSY79 has made important contributions to our understanding of the characteristics, causes and consequences of employment trajectories of US workers. Studies have found that young adults' cognitive and non-cognitive skills affect schooling, wages, employment, work experience and occupational choice.<sup>8</sup> Education and training are consequential; young adults with General Educational Development certificates (GEDs) have labour market outcomes that are more similar to those of high school dropouts than those with a high school diploma,<sup>9</sup> and private sector training plays an important role in determining wages and wage growth of young workers who have less than a college education.<sup>10</sup> The nearly 40-year observation period also enables explorations of macroeconomic influences. For instance, men who graduated from high school during the early 1980s recession had greater difficulty finding a job after graduation and lower wages over the life course, relative to those who graduated during better economic times;<sup>11</sup> similar patterns, yet with more muted effects, also were detected among women.<sup>12</sup>

A second set of studies reveal the importance of early life social and economic influences on health over the life course. Cognitive ability and non-cognitive skills (self-esteem and internal locus of control) in adolescence are positively associated with general health roughly 25 years later, at age 41,<sup>13</sup> and early socioeconomic background and adolescent cognitive ability also account for much of the educational gradient in adults' health behaviours.<sup>14</sup> These multi-wave data enable researchers to identify mechanisms that may link early disadvantage with compromised health at midlife, such as teen childbearing,<sup>15</sup> undesirable employment conditions like irregular work hours<sup>16</sup> and personal bankruptcy.<sup>17</sup> The data also reveal how early life health conditions affect adult socioeconomic prospects. For instance, heavy drinking in high school reduces the average number of years of schooling completed,<sup>18</sup> and obesity suppresses the wages of White women but not of other sociodemographic groups.<sup>19</sup>

Third, studies exploiting the linked NLSY79 and NLSY79 Child and Young Adult data show how maternal characteristics, ranging from health to labour market experiences, affect offspring's health and development. Family structure and income affect routine medical and dental care provision for children,<sup>20</sup> dark and cluttered home environments are linked with the incidence of children's home injuries<sup>21</sup> and the distance between a child's home and hospital affects the use of preventive care.<sup>22</sup> The long observation period and broad range of children's ages also enable researchers to identify correlates of child health at different developmental stages; for instance, young children's health is more vulnerable to

their families' wealth, whereas adolescents' health is more sensitive to current family income.<sup>23</sup> In sum, the NLSY79 data reveal the complex ways in which social statuses, roles and experiences over the life course affect both own and offspring's health, with attention to the cognitive, behavioural and economic pathways that contribute to health disparities.

## What are the main strengths and weaknesses?

### Strengths

The NLSY79 is one of the longest running studies in the USA of a racially, ethnically and economically diverse sample of adults, spanning adolescence through age 60 and beyond. Given the sample's diversity, prospective design, frequent re-interviews and long observation period, the study is particularly well suited for exploring life course influences on health, and evaluating the role of selective survival in health disparities. Roughly half of all participants have a sibling in the sample, allowing explorations of sibling similarities in life course experiences.<sup>24</sup> The NLY79 data can be linked to the companion NLSY79 Child and Young Adult study, fostering analyses of cross-generational influences, such as the impact of maternal employment on offspring health.<sup>25</sup>

Substantively, the NLSY79 offers one of the most comprehensive records to date of the lives of American workers. The study obtains detailed employment histories, with information on weekly employment status, wages, employment gaps, industry, occupation and benefits for every job held since the respondents were in their teens and early 20s. The capacity to combine the employment data with the vast range of information on other topics such as health, cognition, economic well-being and family has allowed researchers to study labour market behaviours in ways otherwise not possible. The NLSY79 also maintains a complete record of family relationships, including histories of all marriages, divorces and many cohabitations, as well as complete child-bearing histories. The study assesses annual income from all sources (collected in every round) and assets (collected every other round), as well as complete histories of place of residence; residential data are available at both the county and census tract levels, allowing researchers to append contextual variables defined at a geographical level.

### Weaknesses

The NLSY79 has several limitations. First, it is primarily a telephone survey, so questions or procedures that rely on face-to-face interaction are difficult to implement, such as anthropometric indicators of health including scale weight,

peak-flow capacity and gait speed. Concerns regarding confidentiality and respondent burden also prevent the collection of biomarker data, and thus health researchers must rely solely on self-reports. Second, the data are not currently linked to administrative files such as Social Security Administration earnings records, due to the difficulty of obtaining respondent consent. The NLS programme is, however, exploring the possibility of this option. Plans are under way to link the data with the National Death Index, which will enhance the study's value as a resource for studying mortality differentials. Third, although the NLSY79 obtains respondent reports of their spouse's demographic and socioeconomic characteristics, the study does not interview spouses/romantic partners directly; thus it has only one partner's perspective on dyadic processes like marital quality, shared health behaviours and couple-level plans for retirement. Finally, the breadth of measurement is a key strength, yet the range of questions asked for any one specific topic at each wave is necessarily limited.

### Can I get hold of the data? Where can I find out more?

The public use data are available free of charge via an online search and extraction site called the NLS Investigator at [<https://www.nlsinfo.org/investigator/pages/login.jsp>]. Extensive online documentation is available at [<http://nlsinfo.org/content/cohorts/NLSY79>]. Researchers can apply to access NLSY79 confidential geocode data at [<https://www.bls.gov/nls/geocodeapp.htm>]. Restricted use NLSY79 zip code and census tract level variables also are accessible at the BLS National Office or a Federal Statistical Research Data Center; see [<https://www.bls.gov/rda/home.htm>].

#### Profile in a nutshell

- The National Longitudinal Survey of Youth 1979 (NLSY79) is a nationally representative sample of US men and women born between 1957 and 1964, who were aged 14 to 22 when first interviewed in 1979.
- The cohort has been interviewed 27 times, annually until 1994 and biennially since then. NLSY79 data span almost 40 years and respondents will be aged 54 to 61 in round 28, to be collected in 2018.
- The female respondents' biological children (born 1970–2011) comprise the companion NLSY79 Child and Young Adult surveys, and have been interviewed 16 times since 1986.

- The study obtains detailed and repeated measures of education, job training, work, health, health behaviours, insurance, substance use, cognition, marriage, children, migration, personality, income and assets, government assistance programme participation and retirement expectations and preparation.
- Expanded health modules were administered at ages 40+, 50+ and 60+, to reflect changing health symptoms, conditions and service use with advancing age.
- Response rates remained above 90% through the first 16 rounds (through 1994), and remained above 80% through the first 24 rounds (through 2010).
- Public use data and documentation are available free of charge at [<https://www.nlsinfo.org/investigator/pages/login.jsp>]. Information about obtaining restricted-use NLSY79 geocode data are available at [<https://www.bls.gov/nls/geocodeapp.htm>].

Enquiries on data access and use should be sent to [[usersvc@chrr.osu.edu](mailto:usersvc@chrr.osu.edu)].

### Funding

The NLSY79 has primarily been funded by the Bureau of Labor Statistics, U.S. Department of Labor. Several topical modules were funded by other governmental agencies. The U.S. Department of Defense supported the military oversample and its interviews from 1979 through 1984, as well as the Armed Services Vocational Aptitude Battery (ASVAB) administered in 1980. The National Center for Research in Vocational Education funded a survey of high schools of NLSY79 respondents and a collection of their high school transcripts. The National Institute of Child Health and Human Development (NICHD) supported the collection of detailed marriage and fertility history data in addition to the entire Child and Young Adult surveys. The National Institute on Alcohol Abuse and Alcoholism and the National Institute on Drug Abuse have funded modules on alcohol and substance use. The views expressed are those of the authors and do not reflect the policies of the BLS or the views of other BLS staff members.

**Conflict of interest:** None declared.

### References

1. Cooksey E. Using the National Longitudinal Surveys of Youth (NLSY) to conduct life course analyses. In: Halfon N, Forrest CB, Lerner RM, Faustman EM (eds). *Handbook of Life Course Health Development*. New York, NY: Springer, 2018.
2. National Longitudinal Survey of Youth. *Retention and Reasons for Non-Interview*. Washington, DC: Bureau of Labor Statistics, 1979.
3. Frankel MR, McWilliams HA, Spencer BD. *National Longitudinal Survey of Labor Force Behavior, Youth Survey*

- (NLS) *Technical Sampling Report August 1983*. 1983. <https://www.nlsinfo.org/sites/nlsinfo.org/files/attachments/130212/NLSY79%20Tech%20Samp%20Rpt.pdf> accessed (20 April 2018, date last accessed).
4. See [https://psidonline.isr.umich.edu/Data/Documentation/pdf\\_doc/psid88w21v1.pdf](https://psidonline.isr.umich.edu/Data/Documentation/pdf_doc/psid88w21v1.pdf) and <https://www.census.gov/programs-surveys/sipp/methodology/organizing-principles/nonresponse.html> (27 April 2018, date last accessed).
  5. MaCurdy T, Mroz T, Gritz RM. An evaluation of the National Longitudinal Survey of Youth. *J Hum Resour* 1998;**33**:345–436.
  6. Aughinbaugh A, Pierret CR, Rothstein DS. Attrition and its implications in the National Longitudinal Survey of Youth 1979. In: *JSM Proceedings*. Alexandria, VA: American Statistical Association, 2017.
  7. Aughinbaugh A, Pierret CR, Rothstein DS. The National Longitudinal Surveys of Youth: research highlights. *Mon Labor Rev* 2015; doi:10.21916/mlr.2015.34.
  8. Heckman JJ, Stixrud J, Urzua S. The effects of cognitive and non-cognitive abilities on labor market outcomes and social behaviors. *J Labor Econ* 2006;**24**:411–82.
  9. Cameron SV, Heckman JJ. The nonequivalence of high school equivalents. *J Labor Econ* 1993;**11**:1–47.
  10. Lynch LM. Private sector training and the earnings of young workers. *Am Econ Rev* 1992;**82**:299–312.
  11. Kahn L. The long-term labor market consequences of graduating from college in a bad economy. *Labour Econ* 2010;**17**: 303–16.
  12. Kondo A. Differential effects of graduating during a recession across gender and race. *IZA J Labor Econ* 2015;**4**:23.
  13. Kaestner R, Callison K. Adolescent cognitive and noncognitive correlates of adult health. *J Hum Cap* 2011;**5**:29–69.
  14. Cutler DM, Lleras-Muney A. Understanding differences in health behaviors by education. *J Health Econ* 2010;**29**:1–28.
  15. Williams K, Sassler S, Frech A, Addo F, Cooksey E. Nonmarital childbearing, union history, and women's health at midlife. *Am Sociol Rev* 2011;**76**:465–86.
  16. Kleiner S, Pavalko EK. Clocking in: the organization of work time and health in the United States. *Soc Forces* 2010;**88**:1463–86.
  17. Addo FR. Seeking relief: bankruptcy and health outcomes of adult women. *SSM Popul Health* 2017;**3**:326–34.
  18. Cook PJ, Moore MJ. Drinking and schooling. *J Health Econ* 1993;**12**:411–29.
  19. Cawley J. The impact of obesity on wages. *J Hum Res* 2004;**39**: 451–74.
  20. Berger LM. Income, family structure, and child maltreatment risk. *Child Youth Serv Rev* 2004;**26**:725–48.
  21. Mott JA. Personal and family predictors of children's medically attended injuries that occurred in the home. *Inj Prev* 1999;**5**:189–93.
  22. Currie J, Reagan PB. Distance to hospital and children's use of preventive care: is being closer better, and for whom? *Econ Inq* 2003;**41**:378–91.
  23. Wolfe JD. The effects of socioeconomic status on child and adolescent physical health: an organization and systematic comparison of measures. *Soc Indic Res* 2015;**123**:39–58.
  24. Bronars SG, Oettinger GS. Estimates of the return to schooling and ability: evidence from sibling data. *Lab Econ* 2006;**13**:19–34.
  25. Miller DP. Maternal work and child overweight and obesity: the importance of timing. *J Fam Econ Issues* 2011;**32**:204–18.