

---

## 2. METHODOLOGY

### Description of National Longitudinal Survey of Youth

To describe drug use and to predict who becomes a heavy drug user, this study uses the Department of Labor's National Longitudinal Survey of Youth–1979 (NLSY79) cohort, representing 33.6 million youth. The NLSY79 is sponsored by the Bureau of Labor Statistics, Department of Labor, to focus on the labor market experiences of adolescents moving into the labor market and adulthood in the United States. For the last 23 years, the NLSY79 has followed a representative sample of about 10,000 youth (aged 14–21 as of December 31, 1978), through their late thirties and early forties by 2000. The survey oversamples blacks and Hispanics, permitting generalizations about these populations. Participants were interviewed annually until 1994 and biennially after 1994.

The NLSY79 has gathered information about the following factors that might influence the entry of adolescents/youth into the labor market:

- Alcohol and substance use;
- Attitudes and aspirations;
- Child care;
- Criminal behavior;
- Family background and demographic characteristics;
- Geographic residence;
- Health conditions, injuries, and insurance coverage;
- Household composition;
- Income and assets;
- Marital and fertility history;
- Military experience;
- Schooling, school records, and aptitude information;
- Training investments; and
- Work and non-work experiences.

The NLSY began in 1979 with a sample of 12,686 youth and continues to have a high response rate (see Exhibit 2.1). In 1980, 12,141 individuals were interviewed with a response rate of 96 percent. By 1990, the response rate was 91 percent. In 1998, the last year included in this study, the response rate was 87 percent, after 17 rounds and 19 years. By 2000, the age of the sample population ranged from 35–42.

This study uses the battery of questions on drug use that were administered in 1984, 1988, 1992, 1994, and 1998. Comparable questions were asked regarding lifetime and past-year use of marijuana, cocaine (between 1984 and 1998), and crack (between 1992 and 1998).

**Exhibit 2.1: NLSY79 Mode and Response Rates, by Year**

YEAR	MODE	SAMPLE SIZE	RESPONSE RATE
1979	Personal/PAPI	12,686	—
1980	Personal/PAPI	12,141	96%
1981	Personal/PAPI	12,195	96%
1982	Personal/PAPI	12,123	96%
1983	Personal/PAPI	12,221	97%
1984	Personal/PAPI	12,069	96%
1985	Personal/PAPI	10,894 <sup>5</sup>	95%
1986	Personal/PAPI	10,655	93%
1987	Telephone/PAPI	10,485	91%
1988	Personal/PAPI	10,465	91%
1989	Personal/PAPI /CAPI	10,605	93%
1990	Personal/PAPI/CAPI	10,436	91%
1991	Personal/PAPI	9,018 <sup>6</sup>	92%
1992	Personal/PAPI	9,015	92%
1993	Personal/CAPI	9,011	92%
1994	Personal/CAPI	8,891	91%
1996	Personal/CAPI	8,636	89%
1998	Personal/CAPI	8,399	87%
2000	Personal/CAPI	8,033	83%

SOURCE: U.S. Department of Labor, Bureau of Labor Statistics, *NLS Handbook, 2000 and 2001*.

<sup>5</sup> A total of 201 military respondents were retained from the original sample of 1,280; 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>6</sup> The 1,643 economically disadvantaged non-black/non-Hispanic male and female members of the supplemental subsample were not eligible for interview in the 1991 survey year. The total number of NLSY civilian and military respondents eligible for interview (including deceased respondents) beginning in 1991 was 9,964.

## Analytic Techniques

This study focuses on two general areas of interest: (1) the movement of individuals into and out of drug use and (2) the prediction of heavy drug use. Each will be presented in turn.

### DESCRIPTION OF MOVEMENT INTO AND OUT OF DRUG USE

In order to depict movement into and out of drug use, we developed “distribution trees” for the five index years during the 15-year period between 1984 and 1998. The distribution trees are based on frequency counts of use starting with 1984 and conditional frequency counts for subsequent years. Frequency counts for 1988 are based on the condition of use or nonuse for 1984, and frequency counts for 1992 are based on condition of use or nonuse for the 1984 and 1988 survey years. Frequency counts for 1994 are in turn based on condition of use or nonuse for the earlier index years. For 1998, the counts are based on condition of use or nonuse for all previous study periods. For example, the distribution of those who used in 1988 is based on use versus nonuse in 1984. For marijuana, cocaine, and a combination of both, two distribution trees were developed—one for those who reported use and another for those who did not report use, starting with the 1984 period.

The trees are also based on crosstabulations (2\*2 tables) that yielded distributions of use and nonuse for two adjacent study periods. Crosstabulations were used to verify the conditional frequency counts developed for the distribution trees. One additional advantage is that these trees permit recalculation of change by combining each of the component cells.

### PREDICTION

The prediction of heavy drug use primarily employs logistic regressions and odds ratios. No claim is made for a direct causal relationship among the variables used in the analysis. Rather, we seek to explain whether certain behaviors or characteristics tend to coincide with the presence of heavy drug use to a greater or lesser degree than do other variables or characteristics.

Logistic regression is often used when the dependent variable is composed of two values, such as the presence and absence of an activity (e.g., was a heavy drug user or was not a heavy drug user). The measure of strength of association is  $r^2$ .

To illustrate the model, let  $H$  denote the measurement of the dichotomous outcome, heavy drug use. Then  $H=0$  if the individual was not a heavy drug user and  $H=1$  if the individual was a heavy drug user. Variables  $W$  through  $Z$  symbolize additional explanatory variables, which may be either dichotomous or continuous. (NOTE: The three explanatory variables are presented for illustrative purposes only; additional explanatory variables do not change the form of the model.). The unified model for dichotomous outcomes is:

$$\text{Logit}(H) = b_0 + b_1*W + b_2*X + b_3*Z$$

Logistic regression generates odds ratio estimates for each predictor. Such estimates are readily interpretable probabilities that indicate how much more likely it is that an outcome would be observed if, all other elements being the same, the predictor occurs compared to when the predictor

does not occur. For example, all other things being equal, an odds ratio would estimate how much more or less likely a man than a woman is to become a heavy drug user. An odds ratio above 1.0 means that the activity becomes more likely, while an odds ratio below 1.0 means that this activity becomes less likely. For continuous variables, we used ordinary least-squares models.

## Variables

Chapter 3 describes drug use for five index years between 1984 and 1998. To portray the movement into and out of drug use, the variables presented are marijuana use, cocaine use, and both marijuana and cocaine use in the past year.

Chapter 4 focuses on predicting drug use. The dependent variable, heavy drug use, is the first focus. There are many ways that the concepts and measures related to heavy drug use have been operationally defined. One operational measure that the Substance Abuse and Mental Health Services Administration (SAMHSA) used in its National Household Survey on Drug Abuse (NHSDA) to define and count individuals who are dependent on drugs is *problematic drug use*. This measure is based on criteria established by the American Psychiatric Association in its *Diagnostic and Statistical Manual of Mental Disorders* versions three (revised) and four (DSM-III-R, DSM-IV). From 1994 to 2000, SAMHSA applied the DSM-IV definition that a person was dependent if s/he met three out of seven dependence criteria<sup>7</sup> (for substances with a withdrawal criterion) or three out of six criteria (for substances without a withdrawal criterion) based on responses to questions in the NHSDA. These questions were revised for the 2000 survey, and questions were added, regarding respondents' potential abuse of drugs (again, based on DSM-IV criteria).

Chronic drug use is a concept used in ONDCP's publication, *What America's Users Spend on Illegal Drugs* (2001). The study defined chronic users by drug use behaviors rather than DSM criteria. The report's authors analyzed the data collected by the National Institute of Justice's Drug Use Forecasting program and operationally defined the concept of chronic drug use as "those who admitted using cocaine or heroin on more than 10 days during the month before being arrested."<sup>8</sup>

Obviously, the differences among these measures of heavy drug use are substantial, ranging from a combination of psychological variables (i.e., the desire for drugs, the difficulty of stopping use, and withdrawal when attempting to end use) as opposed to measures of actual use. The NLSY questions on illicit drugs focus on behaviors regarding drug use, rather than on dependence or abuse. In addition, the categories used in the NLSY to measure level of drug use (e.g., marijuana, cocaine, crack) are high, but capped at 100 or more lifetime uses.<sup>9</sup> That volume of drug use is substantial, but does not contain information on attempts at withdrawal from drug use or on the difficulty of withdrawal.

---

<sup>7</sup> The operational definitions before and after 2000 are described in: Substance Abuse and Mental Health Administration, *National and State Estimates of the Drug Abuse Treatment Gap: 2000 National Household Survey on Drug Abuse*, Appendix A, DHHS, 2002.

<sup>8</sup> W. Rhodes, *ibid.*

<sup>9</sup> There is one exception. In 1984, the measure was capped at 1,000 or more uses.

Eighteen rounds of the NLSY were conducted between 1979 and 1998.<sup>10</sup> Questions on drug use were administered for five of these rounds between 1984 and 1998. Based on data from these years, we developed an operational definition for heavy drug use based on the following criteria:

- Include drugs generally associated with very heavy drug use (cocaine, crack, heroin) as feasible; and
- Select operational definitions that provide a robust number of cases for analysis.

The operational definition of heavy drug use for the analysis in Chapter 4 is both (1) self-reported lifetime use of cocaine or crack on 100 or more occasions, and (2) cocaine or crack use in the year of the interview (it is important to note that not all individuals in the sample responded each year). Heavy cocaine users in any of the five survey years were retained, even if they were nonrespondents to any of the first four rounds (e.g., a “heavy” cocaine user during 1984 was kept in the analyses even if s/he was a nonrespondent after 1984). As presented in Exhibit 2.2, a lenient measure—report of lifetime cocaine or crack use on 100 or more occasions—results in 1,447 heavy drug users. The more stringent measure (e.g., 100 or more lifetime uses of cocaine or crack plus use on 3–4 days per week of cocaine/crack) results in only 127 heavy drug users. We selected a measure that provides enough cases for a robust analysis and also contains a population that used a substantial amount of drugs. This measure—self-reported use of cocaine or crack 100 or more times in one’s lifetime plus cocaine or crack use in the past year— results in 551 individuals defined as heavy drug users.

Other drugs were not included in the operational definition of heavy drug use for several reasons. Heroin and injection drug use were excluded, as the NLSY asked only whether or not the respondent had ever injected drugs. Such knowledge about use without indication of frequency of use did not seem *a priori* heavy drug use. In addition, lifetime use of marijuana was not included as a measure of lifetime heavy drug use, as the operational definition (100 or more times) seemed too low a standard. (NOTE: The range of individuals with 100 or more lifetime uses of marijuana ranges from 2,093 in 1984, to 1,141 in 1992, and 1,330 in 1998. The numbers of individuals who used marijuana in the year prior to each NLSY round are as follows: 3,812 in 1984, 2,213 in 1988, 1,084 in 1992, 1,173 in 1994, and 814 in 1998).

---

<sup>10</sup> The NLSY was also conducted in 2000 and 2002, and another round will be conducted in 2004; none included questions on drug use.

**Exhibit 2.2: Drug Use for Individual Years and Combinations of Year**

	1984	1988	1992	1994	1998	ANY of the 5
Cocaine—Lifetime use 100+ times	240	273	583	315	294	<b>1,335</b>
Cocaine—Used in last year	302	990	323	302	184	<b>2,085</b>
Cocaine—Lifetime use 100+ and used cocaine in past year	191	165	39	93	72	<b>484</b>
Cocaine—Lifetime use 100+ and used cocaine 1–2 days or more per week	82	32	8	26	16	<b>156</b>
Cocaine—Lifetime use 100+ and used cocaine 3–4 days or more per week.	50	20	8	16	12	<b>103</b>
Cocaine or crack—Lifetime use 100+ and used cocaine or crack in past year	191*	165*	71	118	106	<b>551</b>
Cocaine or crack—Lifetime use 100+ and used cocaine/crack 1–2 days or more per week	82*	32*	17	43	29	<b>189</b>
Cocaine or crack—Lifetime use 100+ and used cocaine/crack 3–4 days or more per week	50*	20*	16	24	22	<b>127</b>
Heroin—Lifetime use	119	n/a	126	199	170	<b>416</b>
Injection drug use (without doctor's permission)—Lifetime use	n/a	117**	n/a	206	198	<b>382</b>
Cocaine or crack, Lifetime 100+	240*	273*	678	361	349	<b>1,447</b>
Cocaine or crack 100+ or heroin	329	273*	785	462	430	<b>1,613</b>
Cocaine or crack 100+ or heroin or injection drug use	329	340	785	544	500	<b>1,723</b>
Cocaine or crack, lifetime use 100+, heroin or IV drug use, and used cocaine/crack in past year	222	165	103	153	115	<b>615</b>
Cocaine or crack, Lifetime use 100+ Heroin or IV drug use, and used cocaine/crack 1–2 days or more per week	84	32	32	48	30	<b>209</b>
Cocaine or crack, Lifetime use 100+, Heroin or IV drug use, and used cocaine/crack 3–4 days or more per week	51	20	30	28	23	<b>145</b>

\* Powder cocaine only; \*\*Cocaine injectors

As cocaine and crack are the operational measures, we refined the dependent variable to heavy cocaine use. Cocaine and crack were extensively used in the 1980s and 1990s, the period analyzed in this study.

## Nonresponse Analysis

As with any longitudinal survey, dropouts occur over time. In a study concerned with drug use, it is reasonable to determine if heavy cocaine users (defined above) were more likely to drop out. If, for example, heavy cocaine users were more likely to drop out, the data set may not be representative of all heavy cocaine users who were 14–21 years old on December 31, 1978.

Fortunately, we can test the hypothesis of whether heavy cocaine users were more likely to drop out of the NLSY79 than other youth, using the method of Diggle.<sup>11</sup> Since we have five study periods (1984, 1988, 1992, 1994, and 1998) with data on drug use, this method calls for a separate test on each of the first four time periods.

The first test is whether those who drop out after 1984 (they completed the 1984 interview, but they were not respondents in the 1988, 1992, 1994, or 1998 interviews) are more likely to be heavy cocaine users than those who are respondents in 1984. Of the 12,069 respondents in 1984, 10,700 responded in at least one of 1988, 1992, 1994, and 1998, while 1,369 were nonrespondents to all four as shown in Exhibit 2.3. Among the 10,700 non-dropouts, 1.57 percent satisfied the heavy cocaine user condition in 1984 (100 or more lifetime uses of cocaine or crack, plus use of either cocaine or crack in the last year). It is important that we only use 1984 data to determine the heavy cocaine users, since we only have data beyond 1984 for the non-dropouts. The percentage of heavy cocaine users was 1.57 percent among the non-dropouts and 1.68 percent for the dropouts; this difference was not statistically significant ( $p=.76$ ).

The second test for those who responded in 1988 is very similar. We split this subset into dropouts (nonrespondents in 1992, 1994, and 1998) and nondropouts, and determined the percentages in each who satisfied the heavy cocaine use condition in 1984 or 1988. This test also returned a non-significant  $p$  value ( $p=.43$ ). The third (1992) and fourth (1994) tests were conducted in similar fashion, and the differences were not statistically significant. No test is appropriate for 1998, of course, since this is the most recent year drug-related questions were asked in the NLS79.

In conclusion, NLSY79 dropouts were *not* more likely to have been heavy cocaine users before dropping out of the survey than were non-dropouts.

---

<sup>11</sup> P. J. Diggle, K. Y. Liang, and S. L. Zeger, *The Analysis of Longitudinal Data*, Oxford, England: Oxford University Press, 1994.

**Exhibit 2.3: Four Tests of Non-Random Dropout among Heavy Cocaine Users**

		Dropout?	
		Yes	No
1984	Number of cases	1,369	10,700
	Percent heavy cocaine users	1.68%	1.57%
	Chi-square statistic	0.0942	1 d.f.
	<i>p</i> value	0.76	Not significant
1988	Number of cases	1,704	8,761
	Percent heavy cocaine users	3.11%	2.76%
	Chi-square statistic	0.6309	1 d.f.
	<i>p</i> value	0.43	Not significant
1992	Number of cases	216	8,800
	Percent heavy cocaine users	2.31%	3.41%
	Chi-square statistic	0.7724	1 d.f.
	<i>p</i> value	0.38	Not significant
1994	Number of cases	695	8,196
	Percent heavy cocaine users	5.32%	4.34%
	Chi-square statistic	1.4569	1 d.f.
	<i>p</i> value	0.23	Not significant