# ESTIMATION OF DRUG USE INCIDENCE USING NHSDA

**Robert Johnson, Dean Gerstein, NORC, and Joseph Gfroerer, SAMHSA**
**Robert Johnson, 1350 Conn. Ave. NW, Suite 500, Washington, DC 20036**

**KEY WORDS: Survey design, Survival analysis, Response error**

**1. INTRODUCTION.** Much research points to the importance of understanding patterns of early involvement with drugs. Age of initiation to alcohol and drug use is one of the most powerful predictors of drug consequences and dependence (Turner, W., et al., 1993). Epidemiological studies of school-based populations and clinical research suggest that adolescents who became involved with drugs at earlier ages used drugs more frequently and escalated to higher levels of drug use more quickly (Glantz, 1991).

Despite their importance, patterns and trends in the incidence of drug use, i.e., first use of drugs, have until recently received little attention in research. Epidemiological descriptions of drug abuse in the U.S. in the last three decades primarily have used estimates of prevalence, i.e., reported drug use in a prior reference period, and data on the consequences of drug abuse, such as deaths, emergency room episodes, and treatment admissions. Gfroerer and Brodsky (1992) first presented estimates of trends in the numbers of new users and average ages at first use of marijuana, cocaine, hallucinogens, and heroin based on the combined 1985-1991 NHSDAs.

This paper reviews the methods of data collection and statistical estimation used to estimate rates of drug use incidence in the National Household Survey on Drug Abuse (NHSDA), the primary source of data on the prevalence of drug use in the U.S., and assesses possible biases in these estimated rates. NHSDA is conducted annually by the Substance Abuse and Mental Health Services Administration (SAMHSA) and is a continuing, cross-sectional, personal-interview survey of persons aged 12 and older in the civilian, noninstitutionalized population of the United States. This paper uses data from the 1988, 1990, 1991, 1992, and 1993 NHSDAs. There were approximately 8,000 respondents in each of the 1988 and 1990 NHSDAs and approximately 30,000 in each of the 1991, 1992, and 1993 NHSDAs. The five surveys used essentially the same drug use questions and the 1991-93 surveys oversampled the same six major metropolitan areas.

**2. METHODS OF DATA COLLECTION.** The NHSDA interview takes about an hour to complete and incorporates procedures designed to maximize honest reporting of drug use. Self-administered answer sheets are used by respondents for all drug use questions, except cigarettes, so that responses are not revealed to interviewers. Prior to 1994, the cigarette questions were completed using an interviewer-administered format, which may have resulted in greater underreporting, especially by adolescent respondents (see Section 4). After a questionnaire is completed, the answer sheets are placed in an envelope which is sealed and mailed back to the data collection contractor with no name or address information included.

Data for estimating incidence rates arises from respondents' retrospective reports of their ages at first use of drugs. The NHSDA collects data on first use of about a dozen drugs. Each of the following questionnaire items used to collect data on first use of alcohol, cigarettes, and marijuana has a missing data rate less than 1%:

Cigarettes (interviewer-administered): "About how old were you when you first tried a cigarette?"

Alcohol (self-administered): "About how old were you the first time you had a glass of beer or wine or a drink of liquor, such as whisky, gin. scotch, etc? Do not include childhood sips that you might have had from an older person's drink."

Marijuana (self-administered): "About how old were you the first time you actually used marijuana or hash, even once?"

By using the age respondents first used a specified drug, the age of respondents at the time of the interview, and the calendar date of the interview, estimates can be made of the numbers of initiates and rates of first use in years prior to the survey year (see Section 3.) Since this method of estimating drug use incidence depends upon respondents' retrospective reports of their ages at first use, it is called the "retrospective method."

Gfroerer and Brodsky (1992) compared the retrospective method to two other methods of estimating incidence based on NHSDA questionnaire data. The "difference method" uses data on lifetime prevalence rather than the age at first use: The number of drug initiates in a specified year is estimated by differencing the estimated number of persons who had ever used the drug in the specified year and the estimated number who had ever used in the immediately preceding year. The "direct method" calculates the number of drug initiates using a survey item asking whether the respondent started drug use in the past year. Gfroerer

and Brodsky showed that these two alternative methods yield estimates of trends in incidence that are broadly similar to those resulting from the retrospective method. Yet the alternative estimates are much less precise, since they use data from small subsamples (those who recently used or initiated use of the drug), and they are also biased in their application to the NHSDA, since, given the restriction of NHSDA to respondents 12 and older, the alternative methods do not cover persons who initiated drug use prior to age 12. The remainder of this paper focuses on estimates obtained using the retrospective method.

**3. ESTIMATION.** Since estimated rates of drug use incidence are based on sample survey data rather than on complete data for the population, the rates must be weighted to take into account the complex sample design of the NHSDA. The purpose of weighting the data is to produce statistically unbiased estimates of first drug use in the target population of NHSDA, i.e., the civilian, noninstitutionalized population aged 12 and older in the United States.

The basic sampling weights of each annual NHSDA are equal to the inverses of the probabilities of selection of sample respondents. In other words, the smaller a respondent's chance of entering the sample, the larger the weight of that respondent in the calculation of unbiased estimates for the target population. To obtain the final NHSDA weights, the basic sampling weights are adjusted to reflect household-level and individual-level nonresponse and further adjusted to ensure consistency with U.S. Census population totals. (For technical discussion of NHSDA weighting, see SAMHSA, 1995a, Appendix D.)

These adjustments do not change the interpretation of the weights: The weight of respondent i, say $w_i$, equals the number of persons in the NHSDA target population that are represented by respondent i. For example, if $w_i$ equals 5,000, then the measurements of first drug use of respondent i are assumed to equal the measurements that would have been obtained from 5,000 persons in the U.S. civilian, noninstitutionalized population that are represented by respondent i. The sum of the weights over all respondents estimates the size of the target population, i.e., total persons aged 12 and older in the U.S. civilian, noninstitutionalized population:

$$\sum_i w_i = Size\ of\ population,$$

where the summation is over all NHSDA respondents. Given the near equivalence of sample designs and sample sizes of the 1991, 1992, and 1993 NHSDAs, estimates for the combined 1991-1993 file (e.g., Table 1) were calculated using annual NHSDA weights divided by three.

The key statistic used to gauge first drug use is the "age-specific rate of incidence," also called the "hazard rate." For rare events, each age-specific rate is approximately equal to the probability that an individual who reaches the beginning of a specified year of age without ever using a specified drug will begin using the drug before reaching the end of that year of age. Yet the age-specific rate is a more refined measure than the probability of first use, because it takes into account differences among individuals in the amounts of time they are exposed to risk before first using a drug.

Let RATE (d, a, y) denote the estimated age-specific rate of first use of drug d in age group a during year y. Then RATE (d, a, y) is the ratio of a numerator N, equal to the number of persons who first used drug d at age a in year y (i.e., the number of initiates at age a in year y), and a denominator D, equal to the number of "person-years of exposure" to the risk of first using drug d at age a in year y:

$$RATE(d,a,y) = N/D = \frac{\sum_i w_i\ I_i(d,a,y)}{\sum_i w_i\ e_i(d,a,y)}, \qquad (2)$$

where $w_i$ is the weight of sample respondent i, $I_i$(d, a, y) is a 0-1 variable which takes on the value 1 if and only if respondent i first used drug d at age a in year y, $e_i$(d, a, y) is the estimated number of years of "exposure to risk of first using drug d" of respondent i at age a in year y, and the summations are over all respondents.

The estimates of $I_i$(d, a, y) were calculated using the retrospective estimation procedure developed by Gfroerer and Brodsky (1992). This procedure has two steps: First, using data on reported date of birth, date of interview, and reported age at first use of a drug, the approximate date of first use is calculated for each respondent who reported ever using the drug. Second, the number of new users of the drug in each year is computed by classifying new users by year of first use. The estimates of $e_i$(d, a, y) are based on conventional demographic approximations (e.g., Barclay, 1958): We set $e_i$(d, a, y) equal to 0 either if respondent i did not have a birthday in the age interval a during year y or if

respondent i reported using drug d prior to year y. We set $e_i(d, a, y)$ equal to 1 if respondent i had a birthday in the age interval a during year y, never used drug d prior to year y, and did not initiate use of drug d during year y. We set $e_i(d, a, y)$ equal to .5 if respondent i had a birthday in the age interval a during year y, never used drug d prior to year y, and did initiate use of drug d during year y. The latter approximation assumes that new users of drug d during year y initiated use approximately halfway through their yearlong tenure at age a during year y, so were no longer exposed to risk during the second half of this period. For convenience, estimated age-specific rates are multiplied by 1000, so these rates are typically expressed "per 1000 person-years of exposure."

Table 1 shows estimated rates of first use of alcohol, cigarettes, and marijuana among persons aged 12-17 for years between 1965 and 1990, based on the combined 1991-93 NHSDAs. The trends of the three drugs are quite different: Alcohol increases during the period, cigarettes decline during most of the period, and marijuana increases until 1980 and then declines. The interpretation of the rates is straightforward: For example, the estimated rate of first use of alcohol in 1970 equals 108.2. This means that more than one of ten persons who reached their 12th, 13th, 14th, 15th, 16th, or 17th birthday in 1970 without ever using alcohol used alcohol for the first time before reaching their next birthday. These trends, and results for ten other drug categories, are more fully analyzed in a report to be released by SAMHSA in 1995 (SAMHSA, 1995b).

The results for cigarette initiation in Table 1 do not fully agree with previous research. Cummings et al. (1995) applied methods similar to those presented in this paper to data of adolescent self-respondents to the Tobacco Use Supplements of the 1992 and 1993 Current Population Surveys and concluded that the incidence rate of cigarette smoking among adolescents increased between 1985 and 1990. Cummings et al. (1995) hypothesized that the increase in cigarette smoking initiation was caused by increased promotional expenditures by tobacco companies during the late 1980s.

## 4. SOURCES OF BIAS IN INCIDENCE RATES.

The estimates shown in Table 1 may be subject to three kinds of bias:

- *Bias due to differential mortality.* Some persons who were alive and exposed to the risk of first using drugs during the historical periods presented in Table 1 died before the sampling and data collection of the 1991-93 NHSDAs. The drug-using experiences of these persons were not represented in the results of the 1991-93 NHSDAs. The estimated age-specific rates are biased to the extent that a) individuals who died before 1991-93 made up an appreciable fraction of persons exposed to the risk of first drug use during a given historical reference period and b) the drug initiation patterns of deceased persons during the reference period differed from those of surviving persons. Bias due to differential mortality is likely to be small for historical periods as recent as 1965, because most persons initiate drug use before age 30, and the great majority of such persons were still alive during 1991-93. Based on life tables presented in NCHS (1993), we estimate that about 85% of persons age 30 in 1965 were still alive in 1993.

- *Bias due to memory errors.* There are two principal kinds of retrospective reporting bias due to memory errors, "recall decay" and "forward telescoping." "Recall decay" means the decline in the ability to remember with distance in time from the event or datum to be remembered. "Forward telescoping" means the misperception that past events occurred more recently than they actually did. (See Eisenhower, et al., 1991; Groves, 1989, pp. 420-433; and Holt, et al., 1991 for reviews of recent research on recall decay and telescoping.) In Table 1, recall decay would be likely to result in estimated rates for the earlier part of this period that are biased downward relative to estimated rates for the later part of this period. This is because respondents to the 1991-93 NHSDAs who initiated drug use during the early part of the period must recall events that are more distant in time than respondents who initiated drug use during the later part of the period. In particular, estimates for 1965 and 1970 are based on respondents' abilities to recall events that occurred over 20 years earlier.

- *Bias due to social acceptability and fear of disclosure.* Research on the effects of mode of interview on drug use reporting (Aquilino, 1994; Turner, C. et al., 1992) indicates that drug use reports obtained using interviewer-administered forms result in greater underreporting than reports obtained using self-administered forms. In NHSDA, this kind of bias is expected to primarily affect estimates of cigarette smoking initiation in years after 1987 because a) the cigarette questions were the only drug-use questions that were administered by interviewers rather than self-administered, b) estimates for years after 1987 depend on reports of adolescent respondents, c) purchasing cigarettes is illegal for adolescents in all states, and d) at least one-quarter of NHSDA interviews with adolescents in each year were conducted with someone else in the room at least part of the time (SAMHSA,

1994). The danger of underreporting of cigarette use by adolescents is emphasized in recent DHHS reports (U.S. Department of Health and Human Services, 1994). We expect the interview-administered form of the cigarette smoking questions in NHSDA to exacerbate this form of underreporting.

The three kinds of bias can be expected to have different effects on estimated incidence rates. Bias due to differential mortality affects both the numerators and denominators of rates; similar to nonresponse bias, this kind of bias is small either if relatively few birth cohort members were deceased at the time of the survey or if deceased and nondeceased persons had similar patterns of drug initiation. The survey weights of NHSDA adjust for nonresponse bias, but these weights do not adjust for bias due to differential mortality. Bias due to memory errors or to social acceptability/fear of disclosure affects only the numerators of estimated rates, i.e., estimates of the numbers of drug use initiates.

**5. EVALUATION OF BIAS IN INCIDENCE RATES.** Some evidence for the validity of NHSDA estimates of annual drug use initiates comes from comparing estimates of initiates in the same reference periods computed using NHSDAs conducted in different years. Gfroerer and Brodsky (1992) report that broadly similar estimates of trends in the numbers of new marijuana users in the 1960s and 1970s are obtained using a) 1985-91 NHSDA data and b) NHSDA data obtained prior to 1985. Trends in annual marijuana initiates are also similar when estimated using a) the 1985-91 NHSDAs and b) the 1992-93 NHSDAs (SAMHSA, 1995b).

Table 2 presents the results of an analysis designed to illuminate possible underreporting of cigarette initiation by adolescents. For each of cigarettes, alcohol, and marijuana, the estimated number of initiates age 12-17 in 1987 was estimated using the data from each of five NHSDAs: 1988, 1990, 1991, 1992, and 1993. The respondents whose responses were used in calculating these estimates were aged 13-18 in 1988, 15-20 in 1989, 16-21 in 1990, 17-22 in 1992, and 18-23 in 1993. Our expectation was that, at least for cigarettes, the estimated number of new adolescent users in 1987 would increase regularly from earlier to later surveys, as the proportion of adolescent respondents contributing to the estimation declines. In fact, given large standard errors, Table 2 suggests no clear trend for cigarettes or marijuana. The estimated number of new alcohol users based on the 1988 NHSDA appears unusually large, but one might expect at least one outlier in this analysis of 15 estimated rates.

**6. CONCLUSIONS.** Rates of drug use incidence are critical statistics in monitoring trends in drug use in the U.S. A review of data collection and estimation procedures in the NHSDA suggests three sources of possible bias in such rates: a) differential mortality, b) memory errors, and c) social acceptability/fear of disclosure. The young ages of most first drug users and historical data on U.S. mortality suggest differential mortality is unlikely to be a serious source of bias in rates computed for reference periods less than 30 years ago. Except for the cigarette questions, NHSDA used self-administered answer sheets for all drug questions and other procedures designed to reduce social acceptability bias and respondents' perceived risks of disclosure. (The 1994 NHSDA converted the tobacco section to the self-administered format.) Preliminary analyses comparing estimates for the same reference periods computed using NHSDAs conducted in different years suggest that biases due to memory errors are small for reference periods prior to the 1990s.

**7. REFERENCES.**


Aquilino, W., 1994. "Interview mode effects in drug use surveys," *Public Opinion Quarterly* 54: 362-395.

Barclay, G., 1958. *Techniques of Population Analysis*, New York: Wiley.

Cummings, K., D. Shah, and D. Shopland, 1995. "Trends in smoking initiation among adolescents and young adults- United States, 1980-1989," *Morbidity and Mortality Weekly Report* 44: 521-525.

Eisenhower, D., Mathiowetz, and Morganstein, D., 1991. "Recall error: sources and bias reduction techniques," pp. 127-144 in Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S., eds., *Measurement Error in Surveys*, New York: Wiley.

Gfroerer, J. and M. Brodsky, 1992. "The incidence of illicit drug use in the United States, 1962-1989," *British Journal of Addiction* 87: 1345-1351.

Glantz, M., 1991. "A developmental psychopathology model of drug abuse vulnerability," pp. 389-418 in Glantz, M. and R. Pickens, eds., *Vulnerability to Drug Abuse*, Washington, D.C.: American Psych. Assoc.

Groves, Robert M., 1989. *Survey Errors and Survey Costs*, New York: Wiley.

Holt, D., McDonald, J., and Skinner, C., 1991. "The effect of measurement error on event history analysis,"

p. 665-686 in Biemer, P., Groves, R., Lyberg, L., Mathiowetz, N., and Sudman, S., eds., *Measurement Error in Surveys*, New York: Wiley.

Kandel, D. and Yamaguchi, K., 1993. "From beer to crack: developmental patterns of drug involvement" *Am. Journal of Public Health* 83: 851-855.

National Center for Health Statistics, 1993. *Vital Statistics of the United States, 1989, Volume II-Mortality*, Part A, Rockville, MD.: NCHS.

National Institute on Drug Abuse, 1993. *National Survey Results on Drug Use from the Monitoring the Future Study, 1975-1992: Volume 1, Secondary Students*, Rockville, MD: U.S. Department of Health and Human Services.

Nelson, D., G. Giovino, D. Shopland, P. Mowery, S. Mills, and M. Eriksen, 1995. "Trends in cigarette smoking among US adolescents, 1974 through 1991," *American Journal of Public Health* 85: 34-40.

Pierce, J., L. Lee, and E. Gilpin, 1994. "Smoking initiation by adolescent girls, 1944 through 1988," *Journal of the Am. Medical Association* 271: 608-611.

Substance Abuse and Mental Health Services Administration, 1994. *Preliminary Estimates from the 1993 National Household Survey on Drug Abuse*, Advance Report Number 7, July 1994.

Substance Abuse and Mental Health Services Administration, 1995a. *National Household Survey on Drug Abuse: Main Findings 1992*, DHHS Pub. No. (SMA) 95-3012.

Substance Abuse and Mental Health Services Administration, 1995b. *Trends in the Incidence of Drug Use in the United States, 1919-1992*, DHHS Pub. No. (SMA) 96-3076.

U.S. Department of Health and Human Services, 1991. *Drug Abuse and Drug Abuse Research*, The Third Triennial Report to Congress, DHHS Pub. No. (ADM) 91-1704.

U.S. Department of Health and Human Services, 1994. *Preventing Tobacco Use among Young People: A Report of the Surgeon General*, Atlanta, Ga.: Centers for Disease Control.

Turner, C., Lessler, J., and Devore, J., 1992. "The effects of mode of administration and wording on reporting of drug use," pp. 177-219 in Turner, C., Lessler, J., and Gfroerer, J., eds., *Survey Measurement of Drug Use: Methodological Studies*, Rockville, MD: U.S. Dept. of Health and Human Services, Alcohol, Drug Abuse, and Mental Health Administration.

Turner, W., Cutter, H., Worobec, T., O'Farrell, T., Bayog, R., and Tsuang, M., 1993. "Family history models of alcoholism: age of onset, consequences, and dependence," *Journal of Studies on Alcohol* 54:164-171.

**Table 1. Estimated rates of first use of persons aged 12-17, 1965-90. Per 1000 person-years of exposure. 1991-93 NHSDAs. (SEs in parentheses.)**

| | RATE OF FIRST USE, AGES 12-17 | | |
|---|---|---|---|
| YEAR | Cigarettes | Alcohol | Marijuana |
| 1965 | 114.9 (10.9) | 71.7 (6.7) | 7.0 (1.4) |
| 1970 | 139.4 (8.8) | 108.2 (7.3) | 48.4 (4.6) |
| 1975 | 148.8 (6.5) | 116.7 (5.0) | 74.7 (3.7) |
| 1980 | 126.3 (6.4) | 148.1 (6.6) | 75.6 (3.8) |
| 1985 | 121.6 (6.1) | 143.6 (6.5) | 63.9 (3.6) |
| 1990[a] | 103.6 (4.5) | 165.1 (6.5) | 40.5 (2.5) |

a. Estimated using 1992-1993 data only.

**Table 2. Estimated number in thousands of drug use initiates age 12-17 in 1987. 1988, 1990, 1991, 1992, and 1993 NHSDAs. (SEs in parentheses.)**

| SURVEY YEAR | R's AGE AT INTERVIEW | DRUG | | |
|---|---|---|---|---|
| | | Cigarettes | Alcohol | Marijuana |
| 1988 | 13-18 | 1,577 (173) | 3,024 (282) | 1,298 (213) |
| 1990 | 15-20 | 1,502 (165) | 2,518 (234) | 1,096 (180) |
| 1991 | 16-21 | 1,629 (114) | 2,733 (161) | 1,252 (129) |
| 1992 | 17-22 | 1,482 (103) | 2,145 (127) | 1,141 (118) |
| 1993 | 18-23 | 1,341 (93) | 2,097 (124) | 1,208 (124) |