2000 National Household Survey on Drug Abuse

Statistical Inference

Authors:

Teresa R. Davis
Lisa E. Packer
David C. Heller
James R. Chromy

Project Manager:

Tom Virag - Project Director

Prepared for

Substance Abuse and Mental Health Services Administration
Contract No. 283-98-9008

May 2002

Table of Contents

Statistical Inference

1. Introduction

Starting in 1999 and continuing through 2000, the National Household Survey on Drug Abuse (NHSDA) was implemented as part of a 5-year 50-State sample design to provide national and State estimates of drug use through 2003. A major change to the study protocol was the introduction of computer-assisted interviewing (CAI) methods for both the screening and interviewing of selected respondents.

For the 5-year 50-State design, 8 States were designated as large sample States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas), with samples large enough to support direct State estimates. For the remaining 42 States and the District of Columbia, smaller, but adequate, samples were selected to support State estimates using small area estimation (SAE) techniques.

Using the 50-State design, States were first stratified into a total of 900 field interviewer (FI) regions (48 regions in each large sample State and 12 regions in each small sample State). Within FI regions, adjacent Census blocks were combined to form the first-stage sampling units called "segments." Eight sample segments per FI region were fielded during the 2000 survey year. These sampled segments were allocated equally into four separate samples, one for each 3-month period during the year, so that the survey is essentially continuous in the field. For more detailed information on the sample design, see the 2000 NHSDA sample design report (Bowman, Penne, Chromy, & Odom, 2002).

The final respondent sample of 71,764 persons for the 2000 NHSDA was representative of the U.S. general population (the civilian noninstitutionalized population) aged 12 or older in the year 2000. In addition, State samples were representative of their respective State populations.

2. Sampling Error

The national estimates, along with the associated variance components, were computed using a multiprocedure package called SUDAAN: Software for Statistical Analysis of Correlated Data (RTI, 2001). The final, nonresponse-adjusted, and poststratified analysis weights were used in SUDAAN to compute unbiased design-based drug use estimates. Starting in 2000, all of the

variance estimates were calculated using the SUDAAN option called DESIGN=WR, which is unbiased for linear statistics based on multistage clustered sample designs where the first-stage (primary) sampling units are drawn with replacement. Using only the "with-replacement option" ensured that all sampling error estimates were calculated using the same methodology, which differs from how these estimates were previously computed.

In previous years, a maximum-of-three rule was implemented mainly for quality control purposes in which two additional variance estimates were computed. The second variance was based only on the stratification and unequal weighting effects, and the third was based on no effects, or simple random sampling. The reported variance estimate was then the maximum of these three estimates. This approach was designed specifically for estimates that can be represented as proportions and to ensure that only conservative estimates of sampling error were published. As of November 2000, the decision was made to eliminate the use of the maximum-of-three rule for any future analyses. Section 3 provides a more detailed discussion on the decision to discontinue the maximum-of-three rule.

Because of the nature of stratified-clustering sampling design, key nesting variables were created for use in SUDAAN to capture explicit stratification and to identify clustering. For the 2000 NHSDA, each FI region consisted of its own stratum. Two replicates per year were defined within each variance stratum. The first replicate consisted of those "phasing out" segments (would not be used in the next survey year). The second replicate was made up of those "phasing in" segments (would be fielded again the following year), thus constituting the 50 percent overlap between survey years. Each variance replicate consisted of four segments, one segment for each quarter of data collection**.**

Estimates of means or proportions, $\hat{p}_d$, such as drug use prevalence, take the form of nonlinear statistics where the variances are not capable of being expressed in closed form. Variance estimation for nonlinear statistics in SUDAAN is based on a first-order Taylor series approximation of the deviations of estimates from their expected values (RTI, 2001).

Estimates of domain totals, $\hat{Y}_d$, corresponding to proportion estimates, $\hat{p}_d$, can be estimated as

$$\hat{Y}_d = \hat{N}_d \cdot \hat{p}_d \ ,$$

where

$\hat{N}_d$ = estimated population total for domain *d* and

$\hat{p}_d$ = estimated proportion for domain *d*.

The standard error (SE) for the total estimate is obtained by multiplying the SE of the proportion by $\hat{N}_d$, that is,

$$SE\left(\hat{Y}_d\right) = \hat{N}_d \cdot SE\left(\hat{p}_d\right).$$

This approach is theoretically correct when the domain size estimates, $\hat{N}_d$, are among those forced to Census Bureau population projections through the weight calibration process (Chen, Emrich, Gordek, Singh, & Westlake, 2002). In these cases, $\hat{N}_d$ is clearly not subject to sampling error.

For domain totals, $\hat{Y}_d$, where $\hat{N}_d$ is not fixed, this formula may still provide a good approximation if it can be reasonably assumed that the sampling variation in $\hat{N}_d$ is negligible relative to the sampling variation in $\hat{p}_d$. In most analyses conducted for prior years, this has been a reasonable assumption.

However, for a subset of tables produced from the 2000 data, it was clear that the above approach yielded an underestimate of the variance of a total because $\hat{N}_d$ was subject to considerable variation. In these cases, a different method was used to estimate variances in which a direct variance estimate of the linear statistic that estimates a population total was taken from SUDAAN. Using this method did not materially affect the SE estimates for the corresponding proportions presented in the same sets of tables.

3. Quality Control of Prevalence Rates

For a given variance estimate, the associated design effect is the ratio of the design-based variance estimate over the variance that would have been obtained from a simple random sample of the same size. The NHSDA design involves stratification, clustering, and unequal weighting. Clustering and unequal weighting usually increase the design-based variance (design effect greater than one), but stratification along with effective allocation of the sample can actually decrease the design-based variance relative to what would be obtained using a simple random sample (design effect less than one).

The maximum-of-three rule was developed for sample designs used prior to 1999 when it was generally believed that the combined effects of stratification, clustering, and unequal weighting would always lead to a design effect greater than one. Because there was concern about declaring unwarranted significant results when interpreting data from published reports, using the maximum of the three separate variance estimates provided additional protection against making such errors. As a result of this rule, no published SE estimate ever reflected a

design effect of less than one. This maximum-of-three rule was applied to the 1999 NHSDA reports published through November 2000.

The 50-State sample design implemented in 1999 and continued through 2000 provides very effective geographic stratification and 900 degrees of freedom (*df*) for estimating sampling error for national estimates. An empirical review of the relationships among the three variance estimates and a study of simple variance components lent support to the credibility of some design effects being less than one. The stability of the design-based variance estimates was considered much improved under the new design and larger sample. In addition, the suppression rules used in NHSDA reports would help prevent spurious interpretations of data. As a result, for all 2000 reports, the maximum-of-three rule was discontinued and only the design-based variances and SEs were used.

4. Confidence Intervals

In some NHSDA publications, sampling error was quantified using 95 percent confidence intervals. Because the estimates in the NHSDA are frequently small percentages, the confidence intervals are based on logit transformations. Logit transformations yield asymmetric interval boundaries that are more balanced with respect to the probability that the true value falls below or above the interval boundaries than is the case for standard symmetric confidence intervals for small proportions.

To illustrate the method, let the proportion P_d represent the true prevalence rate for a particular analysis domain "*d*." Then the logit transformation of P_d, commonly referred to as the "log odds," is defined as

$$L = \ln[P_d / (1 - P_d)],$$

where "1n" denotes the natural logarithm.

Letting $\hat{p}_d$ be the estimate of the proportion, the log odds estimate becomes $\hat{L} = \ln[\hat{p}_d / (1 - \hat{p}_d)]$. Then the lower and upper confidence limits of *L* are formed as

$$A = \hat{L} - K\left[\frac{\sqrt{\text{var}(\hat{p}_d)}}{\hat{p}_d(1 - \hat{p}_d)}\right]$$

$$B = \hat{L} + K\left[\frac{\sqrt{\text{var}(\hat{p}_d)}}{\hat{p}_d(1-\hat{p}_d)}\right],$$

where var ($\hat{p}_d$) is the variance estimate of $\hat{p}_d$, the quantity in brackets estimates the SE of $\hat{L}$, and K is the constant chosen to yield a level of confidence (e.g., K = 1.96 for 95 percent confidence limits).

Applying the inverse logit transformation to A and B above yields a confidence interval for $\hat{p}_d$ as follows:

$$\hat{p}_{d,\,lower} = \frac{1}{1+\exp(-A)}$$

$$\hat{p}_{d,\,upper} = \frac{1}{1+\exp(-B)},$$

where "exp" denotes the inverse log transformation. The lower and upper confidence interval endpoints for percentage estimates are obtained by multiplying the lower and upper endpoints of $\hat{p}_d$ by 100.

The confidence interval for the estimated domain total, $\hat{Y}_d$ as estimated by $\hat{Y}_d = \hat{N}_d \cdot \hat{p}_d$, is obtained by multiplying the lower and upper limits of the proportion confidence interval by $\hat{N}_d$. For domain totals $\hat{Y}_d$, where $\hat{N}_d$ is not fixed, the confidence interval approximation assumes that the sampling variation in $\hat{N}_d$ is negligible relative to the error in $\hat{p}_d$.

5. Statistical Significance of Differences

This section describes the methods used to compare prevalence estimates. Customarily, the observed difference between estimates is evaluated in terms of its statistical significance. "Statistical significance" refers to the probability that a difference as large as that observed would occur due to random error in the estimates if there were no difference in the prevalence rates for the population groups being compared. The significance of observed differences is generally reported at the 0.05 and 0.01 levels.

Due to the change in the NHSDA's design, the precision of its estimates of substance use prevalence has improved, but the redesign has also made it difficult to assess long-term trends. It is not considered appropriate to compare 1999 CAI and later CAI estimates with earlier NHSDA estimates to assess trends in substance use because of the major differences between the CAI and the paper-and-pencil interviewing (PAPI) methods. If comparisons such as these are made, it is recommended that they be interpreted with caution.

When comparing 1999 and 2000 prevalence estimates, one can test the null hypothesis (no difference in the 1999 and 2000 prevalence rates) against the alternative hypothesis (there is a difference in prevalence rates) using the standard difference in proportions test expressed as

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\text{var}(\hat{p}_1) + \text{var}(\hat{p}_2) - 2\,\text{cov}(\hat{p}_1, \hat{p}_2)}},$$

where $\hat{p}_1$ = 1999 estimate, $\hat{p}_2$ = 2000 estimate, $\text{var}(\hat{p}_1)$ = variance of 1999 estimate, $\text{var}(\hat{p}_2)$ = variance of 2000 estimate, and $\text{cov}(\hat{p}_1, \hat{p}_2)$ = covariance between $\hat{p}_1$ and $\hat{p}_2$.

Under the null hypothesis, Z is asymptotically distributed as a normal random variable. Calculated values of Z can, therefore, be referred to as the unit normal distribution to determine the corresponding probability level (i.e., p value). Because there is a 50 percent overlap in the sampled segments between the 1999 and 2000 NHSDAs, the covariance term in the formula for Z will, in general, be greater than 0. Estimates of Z along with their p value were calculated using SUDAAN incorporating the analysis weights and accounting for the sample design. A similar procedure and formula for Z are used for estimated totals.

Two non-independent prevalence estimates from the same data year can be compared using the above formula. Using this formula directly and ignoring the covariance term, which is usually small and positive, will result in a somewhat conservative test of hypotheses that will sometimes fail to establish statistical significance when in fact it exists. Comparison estimates computed using SUDAAN would have a more accurate covariance term calculated accounting for the sample design.

6. Incidence Estimates

To assist in the evaluation of trends in initiation of drug use, NHSDA data also were used to generate estimates of drug use incidence, or initiation (i.e., the number of new users during a given year). The 2000 incidence estimates were generated using combined 1999 and 2000 data to

further increase precision and add stability to the trends. Incidence rates measure the rapidity with which new drug users arise and can suggest emerging patterns of drug use.

The measure of incidence is defined as the number of new cases of drug initiation divided by the person time of exposure. For diseases, the incidence rate, *IR*, for a population is defined as the number of new cases of the disease, *N,* divided by the person time, *PT*, of exposure or

$$IR = \frac{N}{PT}.$$

The person time of exposure is measured as the net time that individuals in the population during an observed period of time are at risk of developing the disease. This period of time can be for the full period of the study or for a shorter period. The person time of exposure ends at the time of diagnosis (e.g., Greenberg et al., 1996, pp. 16-19). Similar conventions were followed for defining the incidence of first use of a substance.

Beginning in 1999 and continuing in 2000, the NHSDA questionnaire allowed for collection of year and month of first use for recent initiates. Month, day, and year of birth were also obtained directly or imputed in the process. In addition, the questionnaire call record provided the date of the interview. By imputing a day of first use within the year and month of first use reported or imputed, the key respondent inputs, in terms of exact dates, can be computed. Exposure time can be determined in terms of days and converted to an annual value.

Having exact dates of birth and first use also allowed the determination of person time of exposure during the targeted period, *t*. Let the target time period for measuring incidence be specified in terms of dates. For the period 1998, for example, the specification would consist of

$$t = \left[t_1, t_2\right) = \left[1\ Jan\ 1998, 1\ Jan\ 1999\right),$$

a period that includes January 1, 1998, and all days up to but not including January 1, 1999. The target age group can also be defined by a half-open interval as $a = [a_1, a_2)$. For example, the age group 12 to 17 would be defined by $a = [12,18)$ for youths at least age 12, but not yet age 18.

If person *i* was in age group *a* during period *t*, the time and age interval, $L_{t,a,i}$, can then be determined by the intersection

$$L_{t,a,i} = \left[t_1, t_2\right) \cap \left[DOB_i\ MOB_i\ YOB_i + a_1, DOB_i\ MOB_i\ YOB_i + a_2\right),$$

where the time of birth is defined in terms of day (DOB_i), month (MOB_i), and year (YOB_i). Either this intersection will be empty ($L_{t,a,i} = \theta$), or it was designed by the half-open interval, $L_{t,a,i} = [M_{1,i}, M_{2,i})$, where

$$M_{1,i} = Max\left\{t_1, \left(DOB_i\ MOB_i\ YOB_i + a_1\right)\right\}$$

and

$$M_{2,i} = Min\left\{t_2, \left(DOB_i\ MOB_i\ YOB_i + a_2\right)\right\}.$$

The date of first use, $t_{fu,d,i,}$ is also expressed as an exact date. An incident of first drug *d* use by person *i* in age group *a* occurs in time *t* if $t_{fu,d,i} \in [M_{1,i}, M_{2,i})$. The indicator function, $I_{i(}d,a,t)$, used to count incidents of first use is set to 1 when $t_{fu,d,i} \in [M_{1,i}, M_{2,i})$, and to 0 otherwise. The person time exposure, measured in years and denoted by $e_i(d,a,t)$ for a person *i* of age group *a* depends on the date of first use. If the date of first use precedes the target period $(t_{fu,d,i} < M_{1,i})$, then $e_i(d,a,t) = 0$. If the date of first use occurs after the target period or if person *i* has never used drug *d*, then

$$e_i\left(d,a,t\right) = \frac{M_{2,i} - M_{1,i}}{365}.$$

If the date for first use occurs during the target period, $L_{t,a,i}$, then

$$e_i\left(d,a,t\right) = \frac{t_{fu,d,i} - M_{1,i}}{365}.$$

During leap years, the denominator used to compute person time exposure is set to 366. Note that both $I_i(d,a,t)$ and $e_i(d,a,t)$ are set to 0 if the target period, $L_{t,a,i}$, is empty (i.e., person *i* is not in age group *a* during time *t).* The incidence rate is then estimated as a weighted ratio estimate:

$$IR\left(d,a,t\right) = \frac{\sum_i w_i I_i(d,a,t)}{\sum_i w_i e_i(d,a,t)},$$

where the w_i are the analytic weights.

In previous years, before exact date data were available for computing incidence rates, a person was considered to be of age *a* during the entire time interval *t* , if his/her *a*th birthday

occurred during time interval *t* (generally, a single year). If the person initiated use during the year, the person time exposure was approximated as one-half year for all such persons rather than computing it exactly for each person.

Because of the new methodology, all incidence rates from 1999 and forward are not strictly comparable with prior year estimates. However, because they are based on retrospective reports by survey respondents (as was the case for earlier estimates), they may be subject to some of the same kinds of biases.

Bias resulting from differential mortality occurs because some persons who were alive and exposed to the risk of first drug use in the historical periods shown in the tables died before the 2000 NHSDA was conducted. This type of bias is probably very small. Incidence estimates are also affected by memory errors, including recall decay (tendency to forget events occurring long ago) and forward telescoping (tendency to report that an event occurred more recently than it actually did). These memory errors would both tend to result in estimates for earlier years (i.e., 1960s and 1970s) that are downwardly biased (because of recall decay) and estimates for later years that are upwardly biased (because of telescoping). There is also likely to be some underreporting bias because of social acceptability of drug use behaviors and respondents' fears of disclosure. This is likely to have the greatest impact on recent estimates, which reflect more recent use and reporting by younger respondents. Finally, for drug use that is frequently initiated at age 10 or younger, estimates based on retrospective reports 1 year later underestimate total incidence because 11-year-old children are not sampled by the NHSDA. Prior analyses showed that alcohol and cigarette (any use) incidence estimates could be significantly affected by this. Therefore, there were no 1999 overall estimates (including all ages) made for these drugs.

7. Suppression of Estimates with Low Precision

Direct survey estimates, noted by asterisks (*), are not reported as they are considered to be unreliable due to unacceptably large sampling errors. The criterion used for suppressing all direct survey estimates was based on the relative standard error (RSE), which is defined as the ratio of the standard error (SE) over the estimate.

For proportion estimates $(\hat{p})$ within the range $[0 < \hat{p} < 1]$, rates and corresponding estimated numbers of users were suppressed if

$$\text{RSE}\left[-\ln(\hat{p})\right] > 0.175 \text{ when } \hat{p} \le 0.5$$

or

$$\text{RSE}\left[-\ln(1-\hat{p})\right] > 0.175 \text{ when } \hat{p} > 0.5.$$

Based on a first-order Taylor series approximation to estimate RSE $[-\ln(\hat{p})]$ and RSE $[-\ln(1-\hat{p})]$, the following was used for computational purposes:

$$\frac{\text{SE}(\hat{p})/\hat{p}}{-\ln(\hat{p})} > 0.175 \text{ when } \hat{p} \le 0.5$$

or

$$\frac{SE(\hat{p})/(1-\hat{p})}{-\ln(1-\hat{p})} > 0.175 \text{ when } \hat{p} > 0.5.$$

The separate formulas for $\hat{p} \le 0.5$ and $\hat{p} > 0.5$ produce a symmetric suppression rule; that is, if $\hat{p}$ is suppressed, then so will 1- $\hat{p}$. This is an ad hoc rule that requires an effective sample size in excess of 50. When $0.05 < \hat{p} < 0.95$, the symmetric properties of the rule produce a local maximum effective sample size of 68 at $\hat{p} = 0.5$. Thus, estimates with these values of $\hat{p}$ along with effective sample sizes falling below 68 are suppressed. A local minimum effective sample size of 50 occurs at $\hat{p} = 0.2$ and again at $\hat{p} = 0.8$ within this same interval; so, estimates are suppressed for values of $\hat{p}$ with effective sample sizes below 50.

In previous NHSDAs, these varying sample size restrictions sometimes produced unusual occurrences of suppression for a particular combination of prevalence rates. For example, in some cases, lifetime prevalence rates near $\hat{p} = 0.5$ were suppressed (effective sample size was less than 68 but greater than 50), while not suppressing the corresponding past year or past month estimates near $\hat{p} = 0.2$ (effective sample sizes were greater than 50). To reduce the occurrence of this type of inconsistency, a minimum effective sample size of 68 was added to the suppression criteria in the 2000 NHSDA. As $\hat{p}$ approaches 0.00 or 1.00 outside the interval (0.05, 0.95), the suppression criteria will still require increasingly larger effective sample sizes. For example, if $\hat{p} = 0.01$ and 0.001, the effective sample size must exceed 152 and 684, respectively.

Also new to the 2000 survey is a minimum nominal sample size suppression criteria (*n=100*) that protect against unreliable estimates caused by small design effects and small nominal sample sizes. Prevalence estimates are also suppressed if they are close to 0 or 100 percent (i.e., if $\hat{p} < .00005$ or if $\hat{p} \ge .99995$).

Estimates of other totals (e.g., number of initiates) along with means and rates (both not bounded between 0 and 1) are suppressed if

$$\text{RSE}\left(\hat{p}\right) > 0.5.$$

Additionally, estimates of mean age of first use were suppressed if the sample size was smaller than 10 respondents; also, the estimated incidence rate and number of initiates were suppressed if they rounded to 0.

The suppression criteria for various NHSDA estimates are summarized in Table 1.

Table 1. Summary of 2000 NHSDA Suppression Rules

Estimate	Suppress if:
Prevalence rate, $\hat{p}$, with nominal sample size, *n,* and design effect, *deff*	The estimated prevalence rate, $\hat{p}$, is less than 0.00005 or greater than or equal to 0.99995, or $\frac{\mathrm{SE}(\hat{p})/\hat{p}}{-\ln(\hat{p})} > 0.175$ when $\hat{p} \le 0.5$, or $\frac{\mathrm{SE}(\hat{p})/(1-\hat{p})}{-\ln(1-\hat{p})} > 0.175$ when $\hat{p} > 0.5$, or Effective $n < 68$, or $n < 100$ where Effective $n = \frac{n}{deff}$ Note: The rounding portion of this suppression rule for prevalence rates will produce some estimates that round at one decimal place to 0.0 percent or 100.0 percent but are not suppressed from the tables.
Estimated Number (numerator of $\hat{p}$)	The estimated prevalence rate, $\hat{p}$, is suppressed. Note: In some instances when $\hat{p}$ is not suppressed, the estimated number may appear as a 0 in the tables; this means that the estimate is greater than 0 but less than 500 (estimated numbers are shown in thousands).
Mean age at first use, $\bar{x}$, with nominal sample size, *n*	RSE $(\bar{x}) > 0.5$, or $n < 10$
Incidence rate, $\hat{r}$	Rounds to less than 0.1 per thousand person years of exposure, or RSE $(\hat{r}) > 0.5$
Number of initiates, $\hat{t}$	Rounds to fewer than 1,000 initiates, or RSE $(\hat{t}) > 0.5$

REFERENCES

Bowman, K. R., Penne, M. A., Chromy, J. R., & Odom, D. M. (2002). Sample design report. In *2000 National Household Survey on Drug Abuse: Methodological resource book* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-98-9008). Research Triangle Park, NC: RTI.

Chen, P., Emrich, S., Gordek, H., Singh, A., & Westlake, M. (2002). Sampling weight calibration for the 2000 NHSDA. In *2000 National Household Survey on Drug Abuse: Methodological resource book* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-98-9008). Research Triangle Park, NC: RTI.

Chromy, J. R. (1981). Variance estimators for a sequential sample selection procedure. In D. Krewski, R. Platek, & J.N.K. Rao (Eds.) *Current topics in survey sampling*: *Proceedings of the symposium held May 7-9, 1980, sponsored by the Ottawa Chapter and the Survey Research Methods Section of the American Statistical Association* (pp. 329-347). New York: Academic Press.

Greenberg, R. S., Daniels, S. R., Flanders, W. D., Eley, J. W., & Boring, J. R. (1996). *Medical epidemiology*. Norwalk, CT: Appleton & Lange.

Hansen, M. H., Hurwitz, W. N., & Madow, G. W. (1953). *Sample survey methods and theory*. New York: John Wiley.

Office of Applied Studies. (2001). *Summary of findings from the 2000 National Household Survey on Drug Abuse* (National Household Survey on Drug Abuse Series: H-13, DHHS Publication No. SMA 01-3549). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Research Triangle Institute. (2001). *SUDAAN user's manual: Release 8.0*. Research Triangle Park, NC: Author.

Woodruff, R. S. (1971). A simple method for approximating the variance of a complicated estimate. *Journal of the American Statistical Association*, 66, 411-414.