Population Coverage in the National Survey on Drug Use and Health

Katherine B. Morton, Susan R. Hunter, James R. Chromy, and Peilan C. Martin

RTI International, 3040 Cornwallis Road, PO Box 12194, Research Triangle Park, NC 27709

Key Words: Household Surveys; Response Rates; Coverage Rates

Introduction

The National Survey on Drug Use and Health (NSDUH)[1] is an annual national survey of the civilian, non-institutionalized population aged 12 or older in the 50 United States and the District of Columbia. Chromy et al (1999) examined coverage and response rates for various populations using data from the 1993 through 1998 National Household Surveys on Drug Abuse (NHSDAs). Several design and methodological changes have been implemented since then. Thus, the motivation for this paper is to see if the design and methodological changes had any impact on coverage and response rates as well as to update the prior analysis using data from the 1999 through 2001 NHSDAs and 2002 through 2004 NSDUHs.

History of the NSDUH

Beginning in 1999, the Substance Abuse and Mental Health Services Administration (SAMHSA) implemented major changes in the design of the NSDUH. Prior to 1999, the survey was a national sample of approximately 25,500 persons and was administered through paper and pencil interviewing (PAPI). Beginning in 1999, the annual sample size was increased to 67,500 with minimum sample sizes in each state. The larger sample sizes allowed for estimation at the state level using small area estimation (SAE) techniques. The 1999 survey was also the first to be administered using computer assisted interviewing (CAI) technology.

In 2002, additional changes were implemented. First, the survey was renamed the National Survey on Drug Use and Health (NSDUH). Also beginning in 2002, survey respondents were given a $30 incentive for participation. The 2002 through 2004 surveys could also have been impacted by improved data collection quality control procedures that were introduced in 2001. Finally, the 2000 intercensal projections became available and therefore, the 2002 analysis weights were post-stratified to these totals. All of these changes could impact response and coverage rates, and ultimately drug use estimates. As a result, the 2002 survey became the new benchmark for trend estimation.

---

[1] Prior to 2002, the survey was known as the National Household Survey on Drug Abuse (NHSDA).

NSDUH Sample Design

A five-year sample design was developed for the 1999 through 2001 NHSDAs and 2002 through 2003 NSDUHs. Each state was partitioned into approximately equally sized field interviewer (FI) regions. Within each FI region, area segments were formed by joining adjacent census blocks until each segment contained a minimum of 175 dwelling units. A sample of 96 segments was selected with probability proportional to size within each FI region: 24 to field the five-year sample and 72 "reserve" segments. In order to increase the precision of trend estimates, 50 percent of the segments were carried over from one survey year to the next. The 2004 NSDUH was an extension of the five-year design with approximately 50 percent of the segments being carried over from the 2003 survey and the other half selected from the "reserve" sample.

Within sampled segments, specially trained field staff listed all dwelling units. These lists served as sample frames for the second-stage of selection. At sampled dwelling units (SDUs), FIs entered roster information for all persons 12 or older into a handheld computer which automatically implemented the third stage of selection (persons).

While screening and interviewing in the field, FIs applied the half-open interval rule to pick up any new or missed dwelling units. In summary, the procedure states that any dwelling unit found on the property of an SDU or in the interval between the SDU and the next dwelling unit on the list becomes part of the sample. The procedure minimizes coverage error associated with using "old" listings. If a large number of new or missed dwelling units are found, special procedures are used to select a subsample of the units (Bowman et al, 2005). Typically, we pick up an additional one percent of sample dwelling units through applying the half-open interval rule.

Analysis Weights

The analysis weight is computed as the inverse of the unconditional probability of selection with adjustments for screening and interviewing nonresponse, post-stratification, and extreme weight trimming. The nonresponse adjustments at both the screening and interviewing levels were internal to the sample. The post-stratification adjustment and person-level extreme weight treatment, which adjust for all other frame coverage problems, were based on adjustment of the sample estimates to external intercensal projections by selected

demographics, thought to be very precise in comparison with the survey estimates.[2] For survey years 1999 through 2001, data were post-stratified to intercensal projections based on 1990 census data. From 2002 through 2004, data were post-stratified to intercensal projections based on 2000 census data. Table 1 shows the NSDUH control totals by gender, age group, and race/ethnicity for 1999 through 2004.

Between years 2001 and 2002, when the NSDUH switched from projections based on the 1990 census to projections based on the 2000 census, the total intercensal population estimates increased by 9.5 million. This 9.5 million person increase is a much larger increase than the 2.2 million person increase from 1999 to 2000 and the 2.4 million person increase from 2000 to 2001. When considering the population totals by various subgroups such as age and race, some groups experienced greater population growth than others. For example, the population estimates for Hispanics increased by 4.4 million people between 2001 and 2002, compared with 8 hundred thousand person increases each year from 1999 to 2000 and 2000 to 2001. Another group exhibiting a large increase in population estimates between 2001 and 2002 is the 50 or older age group, which had a 3.4 million person increase. Prior to 2002, annual increases were 1.6 million persons each year from 1999 to 2000 and 2000 to 2001. Since 2002, annual increases have been 1.9 million from 2002 to 2003 and 2.0 million from 2003 to 2004. Thus some populations are growing faster than intercensal projections had anticipated.

Coverage Rate Definitions

Because the data we analyze are impacted by the compounded effects of sampling and nonresponse errors, we examine coverage in a broad sense by looking at screening and interview response rates and coverage rates.[3] We also look at dwelling unit eligibility rates for completeness even though this is not part of coverage. The dwelling unit eligibility rate is defined as the total number of eligible units divided by the total number of selected units. The screening response rate is the number of completed screeners divided by the number of eligible households. At the person level, the interview response

---

[2] Control totals used for post-stratification varied depending on special geographic emphasis of the current survey year, but always included age group (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older), gender, and Hispanic indicator.

[3] While Groves (1989) only includes errors relating to the sampling frame in his definition of *coverage error*, Lessler and Kalsbeek (1992) note that many authors allow the term to include both sampling and nonresponse errors. For example, Kish (1965) defines the term *errors of nonobservation* as the combined effect of these two sources of error.

rate is computed as the number of respondents divided by the number of selected persons. Finally, the coverage rate is computed as the ratio of the weighted survey estimates after adjustment for both screening and interview nonresponse (but before post-stratification and extreme weight trimming) to the external and more precise intercensal estimates.

Results

We first examined overall coverage statistics for the 1999 through 2004 surveys. Table 2 shows the screening response rate, interview response rate, dwelling unit eligibility rate, and coverage rate by year for the 1999 through 2004 surveys. In order to determine the effect of the design and methodological changes on coverage, we also compared the 1999 through 2001 and the 2002 through 2004 medians with the 1993 through 1998 median reported by Chromy et al (1999).

The design and procedural changes had varying effects on the overall coverage statistics. The dwelling unit eligibility rate decreased slightly but was mostly unaffected. Relative to the 1993 through 1998 median, the screening response rate decreased in 1999 through 2001 and decreased slightly more in 2002 through 2004. The 50-state design and the increased sample size put more sample in urban areas and controlled access situations (e.g. gated communities, buildings with buzzers, etc.). Both situations are known to be associated with lower screening response rates. Also, a large number of new interviewers had to be trained which could have contributed to the decrease in screening response rates. The 1999 through 2001 median weighted interview response rate was also slightly lower than the 1993 through 1998 median. Like the screening response rate, the 50-state design made obtaining interview participation more difficult. However, the interview response rate improved substantially in 2002 when respondents began receiving a $30 incentive. Finally, the design and procedural changes had little impact on the estimated frame coverage.

Table 3 examines screening response rates by various segment characteristics. In all years, segments with greater percentages of owner-occupied dwelling units have better screening response rates than areas that are primarily renter-occupied. Similarly, rural areas and largely Hispanic areas have greater screening response rates. Across the board, the 1999 through 2001 and 2002 through 2004 medians are lower than the 1993 through 1998 median screening response rate. The 2002 methodological changes had little impact on the screening response rate.

Next, we looked at interview response rates by demographic group (Table 4). In general, females, 12 to 17 year olds, Hispanics, and non-Hispanic Blacks are the

likeliest to participate. Response rates improved in all subgroups between 2001 and 2002 due to the $30 incentive. The 1999 through 2001 median interview response rate is lower for all groups than it was for 1993 through 1998. The 2002 through 2004 median interview response rate is higher than both the 1993 through 1998 and 1999 through 2001 medians for all groups except for females, persons 50 or older, and Hispanics. The $30 incentive appears to have had the most dramatic effect on the younger age groups.

Finally, we examined coverage rates by gender, age group, and race/ethnicity (Table 5). Coverage improved between 2001 and 2002 for the 12 to 17 and 18 to 25 year old age groups, the two groups that showed the most dramatic increase in interview response rate. Hispanic coverage is consistently high across the years, achieving a 2001 high of 1.0165. It is possible that the NSDUH more accurately estimates the Hispanic population than the intercensal projections. As evidenced by the considerable increase in Hispanic control totals from 2001 to 2002 in Table 1, intercensal projections may under-count the Hispanic population by large numbers. If the intercensal projections are indeed low, NSDUH estimates of the Hispanic population would result in a coverage rate higher than 1.00.

When comparing the 1999 through 2001 median with the 1993 through 1998 median, we show improvements in the median coverage rate for males (increasing from 0.8897 to 0.8975), 12 to 17 year olds (increasing from 0.8512 to 0.8721), 18 to 25 year olds (increasing from 0.8262 to 0.8340), 26 to 34 year olds (increasing from 0.8684 to 0.9165), Hispanics (increasing from 0.9468 to 0.9671), and non-Hispanic Blacks (increasing from 0.8550 to 0.9037). Coverage of 12 to 17 year olds and 18 to 25 year olds continued to increase with the $30 incentive as is shown in the 2002 through 2004 medians.

While the non-Hispanic Black and non-Hispanic White groups have similar 1999 through 2001 and 2002 through 2004 coverage medians with non-Hispanic Black coverage rates slightly higher than non-Hispanic White, the 'Other' group has much larger medians of 0.9761 and 0.9942 in 1999 through 2001 and 2002 through 2004, respectively. Two potential reasons for the consistently high coverage rates in the 'Other' category are (1) the growing multiracial population, and (2) the variability of self-reported race among multiracial individuals, including increased acceptability of specifying more than one race. Thus for these reasons, the multiracial population may be growing faster than the census anticipates.

Conclusions

In general, the NSDUH achieves very good screening and interview response rates and has good coverage of demographic subgroups. Since the 1993 through 1998 NHSDAs, the NSDUH has experienced some variation in these rates due to the implementation of several design and methodological changes. Beginning in 1999, the 50-state design caused moderate decreases in overall dwelling unit eligibility, screening response, and interview response rates while coverage was about the same. Further, the methodological changes that were introduced in the 2002 survey caused increases in interview response and coverage rates for some domains.

As seen in the consistently high coverage rates for Hispanics, the NSDUH either over-counts this population or the intercensal projections are low for this group. Finally, high coverage rates in the 'Other' race category may be due to the growing multiracial population as well as increased acceptability of specifying more than one race.

References

Bowman, K. R., J. R. Chromy, S. R. Hunter, & P. C. Martin (2005, February). 2004 National Survey on Drug Use and Health: Sample design report. In *2004 National Survey on Drug Use and Health: Methodological resource book* (prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-03-9028). Research Triangle Park, NC: RTI International.

Chromy, J. R., K. R. Bowman, C. J. Crump, L. E. Packer, & M. A. Penne (1999). "Population Coverage in the National Household Survey on Drug Abuse." Proceedings of the American Statistical Association, Survey Research Methods Section, American Statistical Association: pp. 576-580.

Groves, R. M. (1989). Survey Errors and Survey Costs. New York: John Wiley & Sons, Inc.

Kish, L. (1965). Survey Sampling. New York: John Wiley & Sons, Inc.

Lessler, J. T. and W. D. Kalsbeek (1992). Nonsampling Error in Surveys. New York: John Wiley & Sons, Inc.

Table 1. NSDUH Control Totals by Gender, Age Group and Race/Ethnicity: Numbers in Thousands, 1999-2004

Domain	1999	2000	2001	2002	2003	2004
Male	106,229	107,344	108,568	113,602	114,985	116,483
Female	114,894	115,935	117,068	121,541	122,697	124,032
12 to 17	23,203	23,368	23,600	24,754	24,995	25,214
18 to 25	28,468	28,984	29,485	31,024	31,728	32,194
26 to 34	33,668	33,010	32,700	35,163	34,961	34,975
35 to 49	63,324	63,882	64,219	65,124	65,031	65,128
50 or older	72,460	74,035	75,632	79,079	80,966	83,004
Hispanic	23,019	23,847	24,662	29,079	29,882	31,030
Non-Hispanic Black	25,240	25,627	25,997	26,809	27,228	27,661
Non-Hispanic White	163,091	163,795	164,605	165,392	166,257	167,051
Other[4]	9,772	10,010	10,371	13,864	14,314	14,773
Total	221,123	223,280	225,636	235,143	237,682	240,515

Table 2. NSDUH Annual Coverage Statistics: 1999-2004

Measure	1999	2000	2001	2002	2003	2004	1993-1998 Median	1999-2001 Median	2002-2004 Median
Screening Response Rate									
Raw	0.9006	0.9299	0.9181	0.9080	0.9102	0.9125	0.9336	0.9181	0.9102
Weighted	0.8963	0.9284	0.9186	0.9072	0.9072	0.9092	0.9343	0.9186	0.9072
Interview Response Rate									
Raw	0.7421	0.7804	0.7681	0.8454	0.8304	0.8266	0.7847	0.7681	0.8304
Weighted	0.6855	0.7393	0.7331	0.7856	0.7739	0.7700	0.7687	0.7331	0.7739
Dwelling Unit Eligibility Rate									
Raw	0.8391	0.8458	0.8427	0.8435	0.8403	0.8413	0.8521	0.8427	0.8413
Weighted	0.8422	0.8491	0.8460	0.8473	0.8416	0.8424	0.8505	0.8460	0.8424
Estimated Frame Coverage	0.9046	0.9025	0.9081	0.8913	0.8884	0.9028	0.9040	0.9046	0.8913

---

[4] Other includes American Indians or Alaska Natives, Native Hawaiians/other Pacific Islanders, Asians, respondents reporting 'other' as their race, and non-Hispanic respondents who report two or more races for years 2002 through 2004. Prior to 2002, the categories Non-Hispanic Black and Non-Hispanic White may include Non-Hispanic persons reporting more than one race since control totals did not include a separate category for multiracial persons during these years.

Table 3. NSDUH Weighted Screening Response Rates by Segment Characteristics: 1999-2004

Domain	1999	2000	2001	2002	2003	2004	1993-1998 Median	1999-2001 Median	2002-2004 Median
50-100% Owner Occupied	0.9047	0.9352	0.9262	0.9140	0.9140	0.9165	0.9401	0.9262	0.9140
10-50% Owner Occupied	0.8773	0.9098	0.9018	0.8914	0.8883	0.8900	0.9215	0.9018	0.8900
< 10% Owner Occupied	0.8551	0.9068	0.8810	0.8692	0.8790	0.8788	0.9082	0.8810	0.8788
Rural	0.9300	0.9455	0.9406	0.9305	0.9348	0.9337	0.9488	0.9406	0.9337
Urban	0.8866	0.9236	0.9126	0.9007	0.8993	0.9021	0.9318	0.9126	0.9007
>50% Hispanic	0.8943	0.9279	0.9356	0.9097	0.9268	0.9269	0.9497	0.9279	0.9268
>50% Non-Hispanic Black	0.8970	0.9246	0.9058	0.9084	0.9064	0.9091	0.9319	0.9058	0.9084
Other[5]	0.9273	0.9428	0.9187	0.9069	0.9060	0.9081	0.9331	0.9273	0.9069

Table 4. NSDUH Weighted Interview Response Rates by Gender, Age Group, and Race/Ethnicity: 1999-2004

Domain	1999	2000	2001	2002	2003	2004	1993-1998 Median	1999-2001 Median	2002-2004 Median
Male	0.6712	0.7268	0.7192	0.7706	0.7572	0.7544	0.7373	0.7192	0.7572
Female	0.6981	0.7509	0.7458	0.7999	0.7896	0.7846	0.7940	0.7458	0.7896
12 to 17	0.7807	0.8258	0.8218	0.8999	0.8957	0.8856	0.8242	0.8218	0.8957
18 to 25	0.7121	0.7734	0.7551	0.8516	0.8347	0.8387	0.7716	0.7551	0.8387
26 to 34	0.6945	0.7492	0.7482	0.7941	0.7869	0.7861	0.7683	0.7482	0.7869
35 to 49	0.6775	0.7389	0.7238	0.7895	0.772	0.7596	0.7622	0.7238	0.772
50 or older	0.6463	0.6953	0.6992	0.7154	0.7077	0.7101	0.7617	0.6953	0.7101
Hispanic	0.7459	0.7795	0.7878	0.8093	0.7955	0.7906	0.8046	0.7795	0.7955
Non-Hispanic Black	0.7039	0.7619	0.7498	0.8224	0.8012	0.8185	0.7855	0.7498	0.8185
Non-Hispanic White	0.6798	0.7339	0.7265	0.7823	0.7721	0.7671	-	0.7265	0.7721
Other[6]	0.5928	0.6731	0.6665	0.705	0.6988	0.6721	-	0.6665	0.6988

[5] Other includes non-Hispanic Whites, American Indians or Alaska Natives, Native Hawaiians/other Pacific Islanders, Asians, respondents reporting 'other' as their race, and non-Hispanic respondents who report two or more races.

[6] Other includes American Indians or Alaska Natives, Native Hawaiians/other Pacific Islanders, Asians, respondents reporting 'other' as their race, and non-Hispanic respondents who report two or more races.

Table 5. NSDUH Frame Coverage Statistics by Gender, Age Group, and Race/Ethnicity: 1999-2004 NSDUH Interview Data

Domain	1999	2000	2001	2002	2003	1993-1998 Median	1999-2001 Median	2002-2004 Median
Male	0.9018	0.8930	0.8975	0.8751	0.8787	0.8897	0.8975	0.8769
Female	0.9073	0.9113	0.9181	0.9065	0.8975	0.9172	0.9113	0.9020
12 to 17	0.8721	0.8580	0.8817	0.9109	0.9190	0.8512	0.8721	0.9150
18 to 25	0.8340	0.8185	0.8345	0.8491	0.8545	0.8262	0.8340	0.8518
26 to 34	0.8909	0.9165	0.9272	0.8656	0.8745	0.8684	0.9165	0.8701
35 to 49	0.9027	0.8992	0.9111	0.9057	0.8973	0.9182	0.9027	0.9015
50 or older	0.9509	0.9460	0.9343	0.9013	0.8912	0.9618	0.9460	0.8963
Hispanic	0.9627	0.9671	1.0165	0.9541	0.9477	0.9468	0.9671	0.9509
Non-Hispanic Black	0.8960	0.9037	0.9287	0.9257	0.8586	0.8550	0.9037	0.8922
Non-Hispanic White	0.8940	0.8884	0.8785	0.8675	0.8736	-	0.8884	0.8706
Other[7]	0.9683	0.9761	1.0696	0.9776	0.9942	-	0.9761	0.9859

[7] Other includes American Indians or Alaska Natives, Native Hawaiians/other Pacific Islanders, Asians, respondents reporting 'other' as their race, and non-Hispanic respondents who report two or more races.