

The Influence of Prior Experiences in Managing Current and Future Risks During Survey Transition Points on the National Survey on Drug Use and Health (NSDUH)

Joe Gfroerer and Jonaki Bose, Substance Abuse and Mental Health Services Administration

Abstract

Since its inception in 1971, the National Survey on Drug Use and Health (NSDUH) has experienced many changes, including the transfer of the survey between federal agencies; changes in government project officers and contractors; modifications to questionnaire content; major changes to the sample design and size; introduction of new modes of data collection and incentive payments to respondents; changes in the oversight and management of field staff; and introduction of new weighting, editing, and imputation methods. Some of these changes have, not surprisingly, resulted in both intended and unintended consequences, in some cases despite best efforts to control and quantify the effects of these changes. This paper uses examples to illustrate how prior experiences have influenced both ongoing practices and the current approach to redesigning the survey.

Key Words: redesign, incentives, context effects, imputation, editing

1. Background and History of the Survey

The National Survey on Drug Use and Health (NSDUH), formerly called the National Household Survey on Drug Abuse (NHSDA), is the Federal Government's primary source of information on the nature and extent of substance use and abuse in the United States. Conducted since 1971, the survey collects data by administering questionnaires to a representative sample of persons aged 12 or older at their places of residence. Data from the survey are used extensively by policymakers and researchers to measure the prevalence and correlates of licit and illicit substance use, to identify and monitor trends in substance use, and to analyze differences in substance use patterns by population subgroups.

Since 1971, the survey has undergone a variety of changes in its sample design as data priorities have changed. During the 1970s and 1980s, it was a relatively small, periodic survey. Conducted every 2 or 3 years, the sample size grew gradually from about 3,000 respondents per survey in the early 1970s to about 9,000 in 1988. In the late 1980s, the nation's cocaine abuse problem became a major concern of the public and of politicians. Congress passed legislation that increased funding for substance abuse data collection and created the White House Office of National Drug Control Policy (ONDCP). ONDCP began producing annual "National Strategies" that used NHSDA data extensively in setting goals and tracking the progress of drug abuse policies and programs. With the increase in funds and greater reliance on NHSDA data by policymakers and researchers, annual fielding of the survey began in 1990, and a significant expansion of the sample began in 1991. The basic national sample size throughout the 1990s was about 18,000 respondents per year.

Throughout the survey's history, interest in particular subpopulations has led to sample design changes and augmentations. Rural areas were oversampled in 1979 and 1994, and the survey oversampled blacks and Hispanics from 1985 through 1998. Supplemental samples of six metropolitan areas were included from 1990 through 1993, and supplemental samples in California and Arizona were added in 1997 and 1998.

For the most part, changes in the data collection methodology prior to 1999 were infrequent and relatively minor. The survey used the same basic methodology from 1971 through 1998: a confidential, anonymous, face-to-face interview conducted in households, employing self-administration of sensitive substance use items. However, some small but important changes were made in the survey procedures that affected survey estimates of substance use prevalence. In 1982, questions on nonmedical use of psychotherapeutic drugs were converted from interviewer-administered to self-administered. Similarly, tobacco questions were shifted to self-administration in 1994. Machine editing procedures were incorporated into the NHSDA data processing for the first time in 1988. In 1994, following extensive research, the NHSDA questionnaire and editing procedures were modified to provide more reliable substance use prevalence estimates.

Methodological research has demonstrated the benefits of audio-computer-assisted self-interviewing (ACASI) in collecting data on sensitive behaviors such as substance use in household surveys. Studies indicate that respondents are more willing to report sensitive behaviors with ACASI than with other modes of data collection. Based on this research, SAMHSA decided in 1995 to initiate development and testing of a computer-assisted interview (CAI), including ACASI, in the NHSDA.

At the same time that the new NHSDA CAI was being developed, a long-standing interest in State-level substance use prevalence data was culminating in legislation that would result in the redesign of the NHSDA sample. With the passage in 1996 of voter initiatives legalizing marijuana use for medical purposes in California and Arizona, and the substantial role of Federal block grant funds given to States for substance abuse prevention and treatment, Congress and the Clinton administration concluded it would be useful to have State-level estimates.

Thus, in 1999, a major redesign of the NHSDA was implemented involving both the sample design and the data collection method of the survey. The national design was changed to a much larger, State-based design with 67,500 respondents per year. The data collection method was changed from a paper-and-pencil interview (PAPI) method to CAI, primarily to improve the quality of NHSDA estimates. Then in 2002, in order to improve response rates and more accurately reflect the focus of the survey, the name of the survey was changed to the National Survey on Drug Use and Health, and an incentive payment of \$30 for all survey respondents was initiated. These two changes, along with enhanced data collection quality control procedures introduced at around the same time, affected survey respondents' reporting of substance use, causing a discontinuity in trend measurement between 2001 and 2002.

2. Current Design of NSDUH

From 2002 through at least 2010, the NSDUH maintained a consistent survey design, which facilitated trend comparisons and pooling of multiple years of data for in-depth analysis. The respondent universe is the civilian, noninstitutionalized population aged 12 years old or older residing within the United States and the District of Columbia. Persons excluded from the universe include active-duty military personnel, persons with no fixed household address (e.g., homeless and/or transient persons not in shelters), and residents of institutional group quarters, such as jails and hospitals. Eight States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) are designated as large sample States with samples of about 3,600 respondents. For the remaining 42 States and the District of Columbia, samples of about 900 persons are selected. Within each

State, samples are equally allocated to three age groups: 12-17, 18-25, and 26 and older. A sample of about 140,000 dwelling units is selected each year. In these sample units (which can be either households or units within group quarters), sample persons are randomly selected using an automated screening procedure programmed in a handheld computer carried by the interviewers. The data collection method used in NSDUH involves in-person interviews with sample persons, incorporating procedures that increase respondents' cooperation and willingness to report honestly about their illicit drug use behavior. Confidentiality is stressed in all written and oral communications with potential respondents. Respondents' names are not collected with the data, and ACASI maximizes privacy and confidentiality.

The NSDUH questionnaire contains a core set of questions critical for basic trend measurement of prevalence estimates that remains in the survey every year and comprises the first part of the interview. Supplemental questions, or modules, that can be revised, dropped, or added from year to year make up the remainder of the interview. The core consists of initial demographic items (which are interviewer-administered) and self-administered questions pertaining to the use of various substances. Complex edits and consistency checks are applied during data processing. For some key variables that still have missing or ambiguous values after editing, statistical imputation is used to replace these values with appropriate response codes.

3. NSDUH's Focus on Trends and Methods Studies

Despite the many methodological changes that have occurred throughout its history, a primary goal of the survey has always been to monitor trends in substance use. This has been a challenge, especially in recent years when the sample size has been large enough to detect very small changes in prevalence rates, including those due to real changes in the population or due to the impact of changes in methods. As a result, the project has maintained a robust program of methodological research, and much has been learned through analyses of the impacts of the survey changes as well as through separate methodological studies designed to investigate specific issues. Results from these experiences and methods studies have provided a valuable base of knowledge which guides the day-to-day management of the survey and decision-making for the redesign of the survey. The next two sections discuss some examples of how prior experiences of methods changes have affected trends and influenced later management of the survey. Two examples involve monitoring systems that have been incorporated into the survey management and three are associated with the planning for a future redesign of the survey.

4. Ongoing Monitoring Systems Resulting from Prior Experiences

4.1 Editing and Imputation Evaluation Report

With the concerns about cocaine abuse and the creation of ONDCP in the late 1980s, NHSDA estimates of cocaine and other drug use received increased attention and scrutiny. In particular, ONDCP chose to use the NHSDA estimate of the number of weekly cocaine users as a key indicator of the prevalence of "hard-core" drug use. But there were important limitations in this NHSDA measure. The NHSDA sample size and design were not optimal for this low-prevalence indicator (less than 0.5 percent), resulting in large relative sampling errors. In addition, reporting bias (underreporting) and undercoverage of "hard-core" drug users were thought to be likely. The release of the

1991 estimates in December 1991 (data collection for the 1991 survey was conducted during January through June), showed an estimated 855,000 weekly cocaine users. This was not a statistically significant increase over the 1990 estimate (662,000), but was nevertheless reported as an increase by ONDCP and the media. Subsequent analysis by project staff revealed anomalies in the imputation of this and other variables, including heroin use, due in part to a slight modification in imputation procedures. A change in the sample design in 1991 (oversampling in six metropolitan areas) had led to a change in the sorting variables used in the hot deck imputation, resulting in a small number of cases with large weights (i.e., persons over age 50) having been inappropriately imputed as drug users. A revised report was issued, with a corrected estimate of 625,000 for weekly cocaine use in 1991. The past year heroin estimate was also revised downward by 46 percent. These revisions generated much attention in the media, a GAO investigation, led to the removal of the Division Director responsible for the survey, and the development of a new monitoring system, called the Editing and Imputation Evaluation Report. This report became an annual product that included a series of tables and analyses to check the impact of editing and imputation on final NHSDA prevalence estimates, by year, to ensure that trends reported were not artifacts of editing or imputation. The report is completed and reviewed by the survey analysts each year prior to the official release of any final estimates.

4.2 Context Effects

Another example of an unplanned methodological effect causing a discontinuity in trends occurred in 2003, with a question context effect. We are referring here to the changes in response patterns to specific questions as a result of a change in the preceding questions. This could be due to changes in the wording of preceding questions, removal of questions, or adding new questions prior to the question of interest. The problem of context effects in survey questionnaires is well known to survey researchers. However, in a large, ongoing survey with hundreds of questions and regular updates, careful monitoring may be necessary to prevent and detect such effects. This issue became evident to NSDUH analysts when initial analyses of the 2003 data revealed that reporting on an item was substantially different from prior years. The question was “How do you feel about adults trying marijuana or hashish once or twice?” Response options were “Neither approve nor disapprove,” “Somewhat disapprove,” and “Strongly disapprove.” In 2002 and before, this question had been preceded by a similar question on attitudes towards use of cigarettes. In 2003, the cigarette question was removed from the questionnaire, changing the context of the marijuana question, apparently affecting the reporting on this question, as seen on table 1.

Table 1: Weighted Percentage Distribution of Responses to 2000-2004 NHSDA/NSDUH Item, “How do you feel about adults trying marijuana or hashish once or twice?”

	2000	2001	2002	2003	2004
Neither approve nor disapprove	30.5	27.9	28.2	43.1	42.3
Somewhat disapprove	13.8	16.4	17.1	13.7	13.5
Strongly disapprove	55.7	55.7	54.7	43.3	44.2

This example convinced the NSDUH team to set up an ongoing system to monitor and evaluate context effects. Whenever questionnaire changes are contemplated, an assessment of potential context effects is done, and options for placement of new questions are evaluated. Potential new questionnaire items were not incorporated into the NSDUH several times based on concerns about context effects. When changes to the

questionnaire are made, the impact on reporting of existing items is analyzed as soon as unweighted partial year data are available. A complete assessment is done with the final fully weighted data and a report on this analysis is done prior to public release of the survey results each year. If context effects are determined to have affected estimates, trends are discussed as appropriate and caveats are documented in appendices that accompany the reporting of results, and in data file codebooks.

5. Prior Experiences Impacting Planning for Survey Redesign

Competing and intertwined with the goal of maintaining trend measurement is the need to periodically update the content and methods of the survey in order to continue to gather relevant and accurate data in a changing environment. As new drugs and new ways of using existing drugs emerge, and as older substances wane in prevalence, change in name, or disappear altogether from the scene, NSDUH survey methods must evolve in order to capture these trends, while remaining flexible to changing budgetary needs. In addition, policy and research focus and priorities also shift, necessitating the addition and removal of questionnaire items. Therefore, NSDUH staff have been planning for a redesign of the survey, possibly to be implemented in 2015. This work began with extensive consultations with principal data users, to learn how they currently use the NSDUH data, and what their current and future needs are. At the same time, a series of methodological studies was initiated, to evaluate current NSDUH sampling, data collection, editing, imputation, and weighting methods, in terms of cost and data quality. New methods such as using an address-based frame for sampling and a new questionnaire structure are being developed and field tested. Various sample design options are being assessed in terms of their impact on cost, precision and burden on field staff. All of this new methodological work is being guided by the results of the various methods studies and experiences gained by project staff in managing the survey for many years. Specific examples of knowledge from past experiences affecting redesign planning are described in the next three sections.

5.1 Impact of Field Staff on Prevalence Rates

The redesign of the survey in 1999 generated a number of interesting methodological findings that have been important considerations in planning the next redesign. One of these considerations was the apparent influence of interviewer experience (on the NHSDA) on the reporting of drug use by respondents. The 1999 redesign involved not only a mode change, from PAPI to CAI, but also a sample design change, from a national sample ($n=25,000$) to a state-based design ($n=70,000$). Furthermore, the 1999 survey design called for an additional 20,000 interviews to enable measurement of the effect of change in mode on key outcome measures and to provide some link between the old and new estimates for trend analyses. Thus, the sample size was 90,000, with 20,000 PAPI interviews and 70,000 CAI interviews. To implement this radically different sample design, the staff of field interviewers had to be increased from about 300 at the end of 1998 to over 1,200 in January 1999. Compounding this difficulty was the requirement to produce state-level estimates, which necessitated a highly unclustered design, with 7,200 first-stage sampling units covering all parts of all 50 states and the District of Columbia. Interviewers were needed in many remote areas that had never been sampled before. As a result, the contractor was not able to ramp up the field staff sufficiently in early 1999, leading to low response rates and a reduced sample size for the 1999 survey. Not surprisingly, a large portion of the expanded interviewing staff was new to survey data collection. While this was not a major concern at first, analysis of the split sample data to

estimate the mode effect revealed an unexpected association between interviewer experience and drug use rates. Respondents interviewed by experienced interviewers were less likely to report drug use than respondents interviewed by new, less experienced interviewers. Of course, interviewer experience was not randomly assigned to sample units, so it was difficult to definitively conclude this, but in extensive modeling with controls for potential confounding variables, the experience effect remained significant. As a result, NSDUH analysts concluded that a valid estimation of the mode effect could not be done with the 1999 split sample, and comparisons between pre-1999 (PAPI) estimates and the CAI data from 1999 and later could not be made.

In subsequent years, followup analyses of the experience effect has shown some decrease over time (possibly due to improved training and monitoring, and attrition of field staff) but still persisted in the NSDUH as late as 2008. This had repercussions for redesign planning. In 2005, when budget cuts were imposed, policymakers looked for options to reduce NSDUH costs without impacting trends. Because a large cut in the sample size would have resulted in a major change in the composition of field staff, and therefore probably the experience level of interviewers, options were limited to small sample cuts and reductions in other activities such as analysis and methods research, which is what eventually occurred. Similarly, for the next major redesign if there is a requirement to preserve trends or accurately measure methods effects during implementation through a split sample design, the sample design may need to in some way control for shifts in interviewer staffing. For example, if the outcome of the redesign planning is that trends must be preserved, and consequently no changes in data collection methods are made, any major change in the sample design (e.g., reverting from a state-based design to just a national design) risks a break in the trend and must be carefully considered.

5.2 Question Structure Trade-offs

A major goal of the NSDUH redesign is to improve and update the questionnaire modules covering misuse of prescription drugs. This is an especially difficult behavior to measure accurately and consistently over time, because of the large number of drugs and the continuing changes in the market. Each year, new drugs and new formulations of old drugs are approved by FDA. Sometimes drugs are discontinued, or switched from prescription to over-the-counter status. NSDUH has attempted to address this by focusing the questions on four major therapeutic classes of drugs (pain relievers, stimulants, tranquilizers, and sedatives), providing photographs of the major drugs in each category and asking brief screener questions about each specific drug, then asking more in-depth questions on usage patterns about only the therapeutic class. This approach assumes that new drugs used by the respondent but not specifically mentioned on the questionnaire would be reported in an "other-specify" question at the end of the drug screener questions. But prior research and actual experience with NSDUH has shown that this can result in serious underreporting bias. In 2006, several questions about the use of recently approved prescription drugs (e.g., Adderall, a stimulant, and Ambien, a sedative) were added to the end of the questionnaire, independent of the main core questions on stimulants and sedatives, which did not include these drugs in screener questions. The result was a five-fold increase in the reporting of lifetime Adderall use and a ten-fold increase in the reporting of lifetime Ambien use. Incorporating these non-core reports of stimulant and sedative use into the estimation and comparing the core-only estimate with the core-plus-noncore estimate revealed substantial bias in SAMHSA's published estimates for stimulants and sedatives (Table 2).

Table 2: Published and Published Plus Noncore Estimates of Nonmedical Use of Stimulants and Sedatives in Lifetime, Past Year, and Past Month among Persons Aged 12 or Older: Percentages, 2006-2009

Drug	2006	2007	2008	2009
<u>Lifetime</u>				
Stimulant Use Published ¹	9.13	8.74	8.49	8.71
Stimulant Use Published Plus Noncore ²	9.81	9.50	9.29	9.75
Sedative Use Published	3.59	3.39	3.56	3.42
Sedative Use Published Plus Noncore ³	4.86	4.99	5.16	5.15
<u>Past Year</u>				
Stimulant Use Published ¹	1.54 ^b	1.21	1.06 ^a	1.22
Stimulant Use Published Plus Noncore ²	1.95	1.64	1.48 ^b	1.77
Sedative Use Published	0.38	0.35	0.25	0.32
Sedative Use Published Plus Noncore ³	1.07	1.08	0.98	1.04
<u>Past Month</u>				
Stimulant Use Published ¹	0.56	0.42	0.36 ^b	0.51
Stimulant Use Published Plus Noncore ²	0.68	0.56 ^a	0.49 ^b	0.68
Sedative Use Published	0.16	0.14	0.09 ^a	0.15
Sedative Use Published Plus Noncore ³	0.32	0.27	0.23 ^a	0.34

*Low precision; no estimate reported.

¹ "Published stimulant" includes noncore methamphetamine as well as published drug use.

² "Published Plus Noncore stimulant" includes published stimulants, noncore methamphetamine, and noncore Adderall.

³ "Published Plus Noncore sedative" includes published sedatives plus noncore Ambien.

^a Difference between estimate and 2009 estimate is statistically significant at the 0.05 level.

^b Difference between estimate and 2009 estimate is statistically significant at the 0.01 level.

These results have significant implications for the redesign of the prescription drug module. First, the new design needs to include a better method for keeping the questionnaire current in terms of capturing the available prescription drug market each year. Second, the correct balance between interview time and the number of drugs included in the screener questions must be determined. The basic conundrum is that while screening for more drugs will result in more accurate estimates for the four therapeutic classes, this multipurpose survey can only devote enough interview time to screen for a limited number of specific drugs.

5.3 Incentives

Another dilemma faced in the redesign, evident from prior experiences with the survey, is the determination of the amount of the incentive payment. The implementation of incentives in 2002 resulted in an increase in the reporting of substance use, so there is concern that any sudden change in the level of the incentive in the redesign could affect trends. Briefly, incentive payments were considered for the NHSDA shortly after the

1999 redesign, due to concerns about declining response rates. Before implementation, a field test was conducted to compare the impact on response rates, costs, and reporting of drug use. The field test, with a sample size of about 4,000, was embedded in the 2001 NSDUH and compared the ongoing no incentive to a \$20 and \$40 incentive payments. The results of the field test showed improved response rates and reduced costs, with no significant impact on prevalence. The incentive effects were most prominent between the \$0 and \$20 options, with marginal effects between the \$20 and \$40 options. Based on these results, SAMHSA applied a \$30 incentive payment starting with the 2002 survey. Other changes to the survey were implemented at the same time, including the name change and increased quality control procedures, which actually began towards the end of 2001 with the intent to reduce the interviewer experience effects discovered in the 1999 data. These changes to the survey in combination caused shifts in the reporting of substance use, with higher proportions of respondents reporting substance use. It was evident that these increases were not reflecting real changes in population behavior, because they were seen in estimates of lifetime prevalence, which should remain similar in the overall population from one year to the next (table 3). It is believed that the incentive payment was the primary methodological change driving these shifts in reporting.

Table 3. Estimates of Lifetime and Past Month Prevalence of Substance Use among Persons aged 12 and older: Percentages

Drug	TIME PERIOD							
	Lifetime				Past Month			
	2000	2001	2002	2003	2000	2001	2002	2003
Any Illicit Drug	38.9	41.7	46.0	46.4	6.3	7.1	8.3	8.2
Marijuana and Hashish	34.2	36.9	40.4	40.6	4.8	5.4	6.2	6.2
Nonmedical Use of Any Psychotherapeutics	14.5	16.0	19.8	20.1	1.7	2.1	2.6	2.7
Any Tobacco	70.5	71.4	73.1	72.7	29.3	29.5	30.4	29.8
Alcohol	81.0	81.7	83.1	83.1	46.6	48.3	51.0	50.1

This experience with the implementation of incentives in 2002 provided important lessons and concerns for the redesign planning. First, it demonstrated the danger in relying too much on relatively small field tests in making decisions about design changes in a very large survey, where small prevalence changes are considered important and are statistically significant. Secondly, the incentive payment of \$30 has been constant since 2002, and its impact has likely eroded. Response rates on the survey increased substantially between 2001 and 2002, but since 2002 they have gradually declined and by 2009 they were back down to about the 2000-2001 levels. So while it seems appropriate to adjust the incentive amount at a minimum to account for inflation, the immediate impact this would have on the response rates and substance use trends is unknown.

6. Conclusions

In summary, prior experience in managing the NSDUH and in conducting various methodological studies have led us to some important conclusions. First, accurate trend measurement with an ongoing cross-sectional survey requires careful monitoring of data collection and estimation procedures to ensure comparability. Caution is needed when

making even small design changes, especially when the sample is very large and sampling errors are small. Small field tests cannot always be relied upon to make decisions about design changes. Secondly, any major redesign of a large ongoing survey is probably going to result in a break in trends. While it may be possible to implement some improvements that have a low probability of disrupting the trend, and it may be feasible to implement a redesign under a split sample design to account for and measure methods effects, there is no guarantee that this will be successful. Promises of trend continuation after a redesign are probably ill advised. The choice often comes down to maintaining the trend with a problematic design and biased estimates, versus improving the survey (including possibly saving on costs) but breaking the trend.

References

Gfroerer, J., Eyerman, J., & Chromy, J. (Eds.). (2002). *Redesigning an ongoing national household survey: Methodological issues* (DHHS Publication No. SMA 03-3768). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies. [Available as a PDF at <http://www.oas.samhsa.gov/nhsda/methods.cfm#Reports>]

Gfroerer, J., Bose, J., Painter, D., Jones, M., Kennet, J. (2009). Redesigning the National Survey on Drug Use and Health. Presented at the annual FCSM conference.

Kennet, J., & Gfroerer, J. (Eds.). (2005). *Evaluating and improving methods used in the National Survey on Drug Use and Health* (DHHS Publication No. SMA 05-4044, Methodology Series M-5). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies. [Available at <http://www.oas.samhsa.gov/nhsda/methods.cfm>]

Office of Applied Studies, Substance Abuse and Mental Health Services Administration. (2003). *Results from the 2002 National Survey on Drug Use and Health: National findings* (DHHS Publication No. SMA 03-38363, NSDUH Series H-22). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies. [Available at <http://www.oas.samhsa.gov>]

Office of Applied Studies, Substance Abuse and Mental Health Services Administration. (2008). *Results from the 2007 National Survey on Drug Use and Health: National findings* (DHHS Publication No. SMA 08-4343, NSDUH Series H-34). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies. [Available at <http://www.oas.samhsa.gov>]

Turner, C. F., Lessler, J. T., & Gfroerer, J. C. (Eds.). (1992). *Survey measurement of drug use: Methodological studies* (DHHS Publication No. ADM 92-1929). Rockville, MD: National Institute on Drug Abuse.

U.S. General Accounting Office (1993, June 25). Drug use measurement: Strengths, limitations, and recommendations for improvement (GAO/PEMD-93-18) Washington, DC: US Government Printing Office.

Wang, K., Baxter, R., and Painter, D. (2005) Modeling Context Effects in the National Survey on Drug Use and Health. Proceedings of the Joint Statistical Meetings