Redesigning the National Survey on Drug Use and Health

Joe Gfroerer, Jonaki Bose, Dicy Painter, Michael Jones, Joel Kennet

Office of Applied Studies, Substance Abuse and Mental Health Services Administration, 1 Choke Cherry Rd, Rockville, MD 20857

joe.gfroerer@samhsa.hhs.gov; jonaki.bose@samhsa.hhs.gov; dicy.painter@samhsa.hhs.gov; michael.jones@samhsa.hhs.gov; joel.kennet@samhsa.hhs.gov

A. Introduction

The Office of Applied Studies (OAS) in the Substance Abuse and Mental Health Services Administration (SAMHSA) conducts the National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years old or older. The survey is used to produce national and state-level estimates of the prevalence of use of illicit drugs, alcohol, and tobacco products as well as measures related to mental health. OAS is planning for a redesign of the survey that will balance three competing goals: 1) maintain valid trend measurement; 2) update and improve the questionnaire and survey methodology; 3) keep costs within expected budget levels. This paper describes the redesign process and the challenges faced in achieving these three goals.

The NSDUH is the federal government's primary source of information on the prevalence and correlates of substance abuse. Its early iterations date back to 1971, when a sample of about 3,000 individuals responded to what was then called the Nationwide Study of Beliefs, Information, and Experiences. In the 1970s and 1980s, the survey was administered every two to three years. As demand and subsequent funding for data on the drug use problem in the United States increased, the survey responded with increases in sample size and complexity, and additional questionnaire domains. Since 1990, the survey has been fielded annually, and in 1999, the current State-based sample design was adopted, with approximately 67,500 individual interviews being conducted each year. The redesign of 1999 also involved a changeover from paper-and-pencil interviewing (PAPI) to the current computer assisted interviewing (CAI) methodology, which combines both computer assisted personal interviewing (CAPI) and audio computer-assisted self interviewing (ACASI). The NSDUH questionnaire covers a variety of topics, including tobacco and alcohol use, misuse of prescription drugs, illicit substance use, substance dependence and abuse, health and mental health, substance use and mental health treatment, environmental issues surrounding treatment, use of healthcare, employment and workplace, income and program participation, and other topics.

Maintaining valid trend measurement is a long standing goal of the NSDUH. In any survey, point estimates are necessarily limited in accuracy by measurement and non-measurement errors. One of the strengths of NSDUH data is its ability to capture trends, due in part to the great deal of effort directed toward minimizing the aforementioned errors. Thus, the need to maintain constancy of methods is of paramount importance. Competing and intertwined with this goal, however, 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 them emerge on the street, 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. Finally, cost is an overarching concern. The costs of updating a survey of NSDUH's magnitude are not small. The processes involved, which are described later in this paper, are time-consuming and require external input at nearly every stage. However, while these costs are at least somewhat appreciable, the cost of continuing to use outdated methods in the presence of newer and more efficient alternatives is seldom recognized, yet equally important to consider.

B. Motivation and planning for redesign

Periodic updates to surveys are necessary in order to keep abreast of changes and developments in the phenomena being studied, and to keep up with the state of the art in survey methodology. Without such updates, a survey will lose its ability to accurately measure the phenomena of interest, and over time will cease to be relevant. To address this, OAS outlined a long-term plan in which the NSDUH would undergo redesign every ten years, based on continuing assessment of data needs and methodological research and development. In April 2005, OAS discussed the proposed long-term plan for redesign with its NSDUH Expert Panel consisting of prominent survey methodologists and substance use and mental health experts. This plan specified a redesign of the survey in 2012, with methodological development and testing to begin immediately. The panel agreed that this was a good approach, as it provided a good balance between the three conflicting goals for the survey mentioned earlier: 1) maintain valid trend measurement, 2) revise the questionnaire to address changing policy and research data needs, and 3) modify the survey methodology to improve the quality of estimates and the efficiency of data collection or processing. Internal discussions within SAMHSA resulted in a delay in the start of the redesign efforts, but work began in

earnest in 2007, driven in large part by concerns about whether the current design could be maintained under expected future budget levels. Cost-cutting measures, including a reduction in the sample size, were prominent in the options considered.

By 2007, a variety of potential design changes to improve the survey had been identified. For example, while the primary focus of the survey had been on illicit drug use among youth and young adults, the need for data on broader issues such as mental health, general health and health care utilization, and drug use among older adults suggested that a shift in the sample design was needed. Problems in the NSDUH questionnaire were also evident, particularly in the prescription drug modules. Historically, the NSDUH has had a 'core' set of questionnaire items that are kept constant from year to year so that the trend measurement process is not confounded by context effects from the placement of new questions before existing ones. Thus, questions added to capture new drugs or to capture more information about existing drugs had to be placed later on in the questionnaire. The addition of new items at the end of the questionnaire forces respondents to jump from topic to topic, returning to some that were asked about previously. This is clearly not the optimal way to aid in recall and reduce respondent burden. Nevertheless, the data from these new items at the end of the interview showed that the current NSDUH core items were outdated, resulting in significant bias.

The collection of data on methamphetamine use is an example of how additional items have been added to the end of the questionnaire in order to improve measurement of particular drug use behaviors. Although methamphetamine is available in prescription form (e.g., Desoxyn®), most of the methamphetamine used nonmedically in the United States is produced by clandestine laboratories within the United States or abroad rather than by the legitimate pharmaceutical industry. Until 2005, the NSDUH only asked about methamphetamine use in the core section of the questionnaire in the context of prescription stimulant drugs that could be used for legitimate medical reasons as well as for nonmedical ones. The concern was that some methamphetamine users would fail to report the use of the drug when it is presented in the context of prescription drugs leading to the underestimation of the true prevalence of methamphetamine misuse in the United States. To address these concerns, new questions were added after the core section, to the noncore special drugs module in the 2005 NSDUH to capture information from respondents who may have used methamphetamine but did not think of it as a prescription drug and therefore did not report use in the core stimulants module. Inclusion of noncore data for estimating methamphetamine use had a significant effect on prevalence estimates of nonmedical use of methamphetamine compared with the corresponding estimates based on core data alone (1.1 percent vs. 0.7 percent respectively for past year use) (Ruppenkamp et al, 2006).

The need for new methamphetamine questions arose in order to accurately capture its use as a street drug. The NSDUH also has to account for the abuse of new prescription drugs. For example, questions regarding the use of Ambien, a new prescription sedative, were added to the noncore section of the 2006 survey. Because these new questions about lifetime and most recent use of Ambien were asked after the respondent had answered all of the questions in the core drug modules (i.e., tobacco through sedatives), data from these new questions did not affect how respondents answered the core questions. Consequently, the measures and data from the core drug sections in 2006 were comparable with data from these sections in 2005 and prior years, for the purpose of measuring trends. However, it turned out that many respondents who had not reported use in the core sedative module did report using Ambien when asked the new questions later in the interview. Inclusion of noncore data for Ambien had notable effects on past year prevalence estimates of nonmedical use of sedatives compared with the corresponding estimates based on core data alone (1.1 percent vs. 0.4 percent respectively) (OAS, 2008).

Other problems with the existing NSDUH questionnaire were evident. When the questionnaire was redesigned and converted to CAI in 1999, there were major concerns about the length of the ACASI portion of the interview, since no large-scale survey had ever fielded a 30-minute ACASI instrument. However, NSDUH has shown that respondents have no problem with completing ACASI questionnaires of this length. Several items being asked in the interviewer-administered sections are candidates for inclusion in the ACASI. These include questions about income, skipping work or school, and other sensitive topics that would best be answered in private. There is also a problem in that the street names for several drugs have undergone changes. Respondents might incorrectly report non-use of substances with street names other than those with which they were familiar. Items with unbalanced sets of response options and other specific, well-known deficiencies had made their way into the questionnaire over time for a number of reasons, and were in need of rewriting.

In the fall of 2007, OAS started establishing steps and a schedule for the redesign including the methodological work that would precede the redesign. The assessment of current and future data needs was established as the first priority. In addition, several major areas of investigation were identified as priority areas for methodological research. These were (1) sampling, (2) improving response rates, (3) questionnaire, and (4) estimation. Within these broad areas of focus, a series of methods studies and consultations were proposed, discussed, and refined by OAS in conjunction with RTI International, the contractor conducting the NSDUH. Work plans and budgets were developed. Importantly, any studies involving data collection had to

be addressed with careful scheduling, to coordinate results with other studies and to plan for OMB and IRB clearances. Early in the planning, it became clear that the 2-year delay in the start of methods development and testing had severely limited the ability to conduct the work necessary to field the redesign in 2012. Additional funding was provided to ensure that the current design could be maintained beyond 2011, and in 2009 SAMSHA decided to delay the implementation until 2013, with a possibility of extending it an additional year.

The following sections provide summaries of the various activities and methods studies conducted in conjunction with the NSDUH redesign. Since all of this new methods research was planned and conducted in the context of prior research relevant to NSDUH, a summary of key previous findings is therefore included as well

C. Assessing data needs

One of the challenges faced by statistical agencies that make tabular and micro data publicly available without any restrictions is the difficulty knowing exactly what data are used, how they are being used and who are using the data. As a preliminary step to defining the sample design and questionnaire contents, a comprehensive outreach was conducted to known key NSDUH data users to obtain feedback on data needs and to solicit specific recommendations on the redesign of NSDUH, particularly the questionnaire content and populations of inferential interest.

Feedback was requested from within SAMHSA, other Federal agencies, state government agencies, selected outside researchers and data users and subject-matter experts focusing on specific topics such as mental health treatment, prescription drug abuse and tobacco use. In early 2008, in-person presentations presenting the OAS redesign plans were made to other offices within SAMHSA and an e-mail outreach was sent to federal, academic and other researchers. Information on the current questionnaire content, how it has evolved since the 1999 redesign and new topics under consideration was provided to each contact. All data users were asked for recommendations for changes to the survey based on their data needs. Responses and recommendations received were summarized and incorporated in to the redesign methodological work for further study and consideration.

In May of 2008, a survey specifically focusing on state governmental needs and data users was conducted to assess how states use the substance abuse and mental health data that are collected in the NSDUH and to solicit suggestions for changes to the survey that could increase its utility. Contacts (one or more in each state) in all 50 states and the District of Columbia were identified by SAMHSA's Center for Substance Abuse Treatment (CSAT) through their work with state substance use agencies and were asked to participate in the survey via a self-administered, web-based questionnaire. The questionnaire asked whether or not the States used the NSDUH data; if so, in what ways; and if not, why not. Respondents were also asked to provide recommendations for additional topics or design changes that would make the NSDUH data more useful for their State. The survey was a success— responses were received from all 50 states and the District of Columbia. All 50 states reported being aware of the NSDUH data and 49 states reported using the data in some capacity. The most commonly reported uses of NSDUH data were informing the public and reporting to state epidemiology workgroups. Thirty-seven states reported using NSDUH data for developing policy and legislation, and 22 states reported using the data for allocating funds. Recommendations from users mainly centered around new survey estimates and new data products.

As much as possible, OAS would like the new NSDUH design to be driven by data needs. This feedback will guide OAS as it proceeds with the methodological development of the new survey design as well as future budget formulations.

D. Prior research studies relevant for the NSDUH redesign

The following are some previous methods research results that will be used in the redesign planning and decision-making.

Validity Study – In 2000 and 2001, NSDUH collected hair and/or urine from a subsample of 4,000 respondents age 12 to 25, to check the validity of self-reports by testing the biological specimens for drugs. The study showed that high cooperation rates (89 percent of those selected provided at least one specimen) can be achieved, and found evidence of underreporting of some drugs. However, interpretation was not straightforward because of uncertainties about time periods covered by tests, and inability to match the tests with the NSDUH reference periods in the self-report data. This study provided information on the feasibility, challenges and limitations of collecting biological specimen data in the NSDUH.

Reliability Study – A reliability study was conducted as part of the 2006 NSDUH to assess the reliability of responses to the NSDUH questionnaire. An interview/reinterview method was employed in which 3,136 individuals were interviewed on two occasions during 2006 generally 5 to 15 days apart; the initial interviews in the reliability study were a subset of the main study interviews. The reliability of the responses was assessed by comparing the responses of the first interview with the responses from the reinterview. Results from this study are being used in the questionnaire redesign process to identify questions with poor reliability.

Interviewer Effects – Several analyses using data prior to 2001 have shown that experienced NSDUH interviewers obtained lower reporting of drug use than inexperienced interviewers. This raises concerns about the impact of a major sample redesign or changes in the sample size that would alter the composition of the field staff and therefore affect prevalence rates and trends. One of the goals of the redesign process is to identify a flexible survey design that can scale up or down, or increase sample for particular geographic areas or subsamples without affecting trends. Understanding the relationship between field staff composition, including experience and key outcome measures is crucial for this.

Mode effects – Numerous studies done as a part of NSDUH and also by outside researchers have shown that mode effects exist for substance use and mental health variables. The effect has been shown to vary by age and other demographic variables, and for different measures, and is related to the privacy of the mode and setting, and the sensitivity of the behavior. As NSDUH considers the expansion into multi-mode data collection (the mental health surveillance study within the NSDUH already collects data over the telephone) it is important to understand mode effects on its estimates.

Incentives impact – NSDUH began offering a $30 incentive for completion of the interview in 2002. Analysis showed that in addition to increasing response rates, the incentive also resulted in increased reporting of substance use, causing a break in NSDUH trend lines. Given the cost of conducting a full-scale experiment to determine the appropriate incentive amount for the redesigned survey, OAS is planning to simply apply an inflation adjustment to the $30 level that was determined based on a field test done in 2001 and used in the survey since 2002. It is anticipated that this will result in an incentive of approximately $40 by 2013 or 2014.

Mental Health Surveillance – SAMHSA decided to expand the mental health component of NSDUH to provide estimates of serious mental illness (SMI) and data related to suicide. In 2008, new questions were added to the survey and a follow-up interview component was implemented, in which a random subsample (n=1,500 in 2008; n=500 in subsequent years) of respondents is administered (by telephone) the Structured Clinical Interview for DSM-IV-TR Axis I Disorders Non-Patient Edition (SCID) to determine psychiatric diagnoses. These 1500 cases were used to create a model that predicted SMI. The model inputs are variables collected for all adults in the national survey and was calibrated against data collected in this special study. The success in creating a model to predict SMI for adults in the NSDUH will affect what questionnaire items will be retained in the redesigned NSDUH questionnaire.

NSDUH Review – In 2007, SAMHSA commissioned an independent (i.e., managed by SAMHSA's Center for Substance Abuse Treatment (CSAT)) review of NSDUH methods to assess various cost-cutting options (SAMHSA, 2007). The study by Westat concluded that the only way to achieve significant cost savings is to substantially redesign the sample and cut the sample size, with a high risk of disrupting the trend lines.

E. New research related to sampling

1. United States Postal Service (USPS) frame study

The current NSDUH sample frame development work includes costly counting and listing within each sampled segment nationwide. This study aims to develop and test an alternative sampling frame largely based on residential mailing address lists – a much more cost-effective approach than traditional field enumeration (counting and listing). Since a frame based exclusively on locatable mailing addresses will under-cover portions of the NSDUH target population, the alternative sampling frame will be based on a combination of locatable mailing addresses, field enumeration, and a Half-Open Interval (HOI) frame-linking procedure. As part of the exploratory work, a 200-segment field test was conducted that evaluated features between the two frames. This field test, conducted in spring 2009, was intended to answer questions regarding the expected coverage of the alternative frame nationally and by region of the country, urban/rural areas, group quarters, and areas on and near military installations. Specifically, the study sought to determine whether there are differences in prevalence estimates between persons picked up by the alternative frame and those picked up by a 100% field-enumerated frame, whether it is possible to establish a mailing list coverage threshold below which segments should be field enumerated,

and what time and cost savings are actually realized by the use of a USPS address frame. Based on the results, decisions will be made on whether and how to proceed with the implementation of a USPS address-based frame. Additional field testing is also planned prior to full implementation.

2. Sample options study

This research examines the feasibility and impacts of several potential sample design changes in terms of cost versus potential changes in precision. Some of the different design issues being studied include determining the optimal number and size of clusters (PSUs) for the NSDUH, evaluating the pros and cons of conducting the survey every other year, and determining whether conducting the survey every year with one-half the sample is a viable alternative. There is also an investigation of the pros and cons of utilizing a more continuous sample in the NSDUH rather than the quarterly design currently being used. Current estimates and variances for various age groups and demographic subgroups are being examined to determine which groups may be targeted for oversampling to improve overall precision and to address changing data priorities. The study is evaluating the potential effects of expanding the NSDUH target population to include children under the age of 12, as related to sample design, nonresponse (parental consent), and data quality are also taken into consideration.

As discussed before, one of the goals of the redesign is to implement a sample design that allows for flexibility, such as drawing additional primary sampling units (PSUs) to allow for supplemental studies, adjusting sample size nationwide to reflect budget changes, or increasing sample for particular subgroups or geographic areas in response to specific needs. To that effect, two options being examined are increasing the number of state sampling regions, and the adoption of a design that enables transition to a national design from the current 50-state design without affecting trends.

All of these factors are being considered collectively, since decisions in one area can narrow the options in others. For example, whether future NSDUHs are designed to produce state or national estimates will affect factors such as cluster sizes, variances, field staff size and composition, travel distances, costs, etc.

3. Investigation the use of ACS data for sample design and other applications

Traditionally, the NSDUH sample has been designed using decennial Census data supplemented with revised population projections. However, 2010 Census data will not be available in time to use for the redesigned 2013 NSDUH and the Census will no longer use the long form, causing some demographic and socio-economic variables to be missing. The purpose of this work was to investigate whether the American Community Survey (ACS) data fit specific requirements. The ACS data must be timely enough to be used for sample design, must include the proper demographic and socio-economic variables for weighting, and must be collected at a small enough geographic domain for small area estimation.

F. New studies related to improving response rates

Initial discussions included a number of proposals to improve response rates, including increased outreach, the use of telephone screening, and optimizing callback models. Redesign efforts are currently focusing on developing improved and updated contact materials to increase respondent cooperation. This study will utilize literature reviews, expert review, and focus groups to identify changes that may help to improve cooperation rates. NSDUH contact materials have to be designed to provide adequate information to encourage respondent participation without potentially frightening them off with the subject matter or raising concerns about the requirement to use a computer. This task is examining updates to the *lead letter* including links to the respondent website, endorsements, the size and placement or the logos, and language used. The *envelope* in which the lead letter is mailed is also a candidate for updating and factors such as color, size, and use of a government logo are being considered. The *question and answer brochure* is also being revised based on past focus groups and experts' suggestions to provide relevant information that is easily understood and will motivate a respondent to participate.

G. New studies assessing NSDUH estimation methods

1. NSDUH weighting assessment

The purpose of this investigation is to improve the survey estimates by reducing sampling error and nonresponse and coverage bias. A second goal is to shorten the weighting process through changes in the current weight calibration procedures. The weighting assessment will examine a number of dimensions such as looking at the possibility of using finer age and race categories during the nonresponse and postratification adjustments, the removal of particular postratification level adjustment steps, and whether it is possible to reduce the number of control variables currently used during the poststratification process. Preliminary analyses indicate that the use of additional age and race categories increases the quality of weighting and that there are no variables recommended for dropping in the poststratification process.

This assessment is also examining the incorporation of paradata on interviewer contact and interview attempts (i.e., callback data) for both respondents and nonrespondents into the weighting. The effectiveness of the current nonresponse adjustment step and the callback model alternative are being compared in terms of reduction in bias, software efficiency, complexity of modeling, runtime, cost, impact on the annual data processing schedule, effect on field staff and consistency in measures over a period of time. Preliminary analyses show that while these callback models may possibly be used to ascertain the presence of potential nonresponse bias, they do not show sufficient increase in data quality when used by themselves for weighting adjustments. Therefore, the benefits of simply incorporating these callback-related paradata into the existing weighting models, instead of using separate call back models is being examined.

2. Predictive Mean Neighborhood (PMN) imputation evaluation

Since 1999, the Predictive Mean Neighborhoods (PMN) procedure has been used to impute missing values for many of the analytical variables in NSDUH (Aldworth, et al. 2005), but it has not been formally evaluated. One key feature of the PMN procedure is that it imputes the use of different drugs step by step, incorporating the results of the imputation of one drug into the imputation of the next drug following a substantively predefined order. This study evaluates the PMN against other simpler methods, and whether sequentially imputing drugs produces different or better estimates. Based on preliminary results, it appears that using IVEware for imputation and the computation imputation-induced variance in the NSDUH will not be possible.

3. Small Area Estimation (SAE) evaluation

The NSDUH uses small area estimation (SAE) procedures in its state and substate-level estimates (Hughes et al., 2009). In general state-level estimates use two years of pooled data, while substate estimates use three years of pooled data. The purpose of this study is to evaluate the quality of current substate estimates, determine the most efficient methods for estimating trends, and compare model-based vs. design based estimation options for different analyses.

H. Developing a new questionnaire

1. Questionnaire structure study

Survey conditioning may occur when earlier questions in a survey affect a respondent's behavior when he or she answers questions later in the survey. In the current *interleafed* format of the NSDUH questionnaire, respondents may learn that answering "yes" to an initial gate item leads to follow-up questions, which might make them more likely to answer "no" to subsequent gate items. The purpose of this study is to examine an *ensemble* format that will present all of the core drug use gate questions before any follow-up questions are asked with the hope that it will produce more accurate answers especially for the drugs asked later on. Changes to the placement and wording of other questions are also being tested.

2. Electronic pillcards and reference date calendar development

The current NSDUH instrument provides two forms of visual cues: a reference date calendar and pillcards. Currently, a paper copy of the reference date calendar is provided respondents at the beginning of the interview. Respondents are prompted by the computer to request the pillcards from the field interviewer (FI) during certain sections of the interview. Respondents sometimes ignore or overlook this prompt, so they lose the opportunity to use this important recall tool. The purpose of this

study is to develop and integrate electronic pillcards and an electronic reference date calendar into the NSDUH CAI instrument. These items will appear on the computer screen as appropriate, ensuring that all respondents view the recall tools. This may potentially increase accuracy in reporting.

3. Debriefing questions and persuasive statements assessment

Interviewer and respondent debriefing questions are often utilized in questionnaires to assess the potential for measurement error. Honesty appeals, or statements emphasizing the importance of accurate reporting, are sometimes included in surveys with the expectation that they will increase the accuracy of respondents' answers. NSDUH currently includes only FI debriefing questions and does not include a strong honesty appeal. This study consisted of a review of relevant surveys both with and without debriefing questions to determine if other surveys are using debriefing items as data quality indicators. Based on the results of the review, a candidate list of FI and respondent debriefing questions have been developed. A short set of FI and respondent debriefing questions, and a short version of an honesty appeal will be included in the questionnaire field test.

4. Prescription drug module redesign

Misuse of prescription psychotherapeutic drugs is second only to marijuana as the nation's most prevalent illicit drug use issue. Measurement of prescription drug misuse is complicated by factors such as frequent changes in the pharmaceuticals available on the market and shifts in physicians' prescribing practices. NSDUH defines "nonmedical use" as use without a prescription or only for the experience or feeling the drug caused. Most questions cover four therapeutic classes (prescription pain relievers, tranquilizers, stimulants, and sedatives), but respondents also are asked about lifetime nonmedical use of specific drugs within these classes. This study is designed to evaluate how best to define and operationalize "nonmedical use" in NSDUH; determine what therapeutic classes of prescription drugs should be covered in NSDUH, and how they should be described to respondents; identify the specific prescription drugs to be included in the questionnaire, and determine the periods of use that are of most interest; and develop a method of updating the list of specific prescription drugs asked about while maintaining continuity for trend analysis.

5. Text-to-Speech Software (TSS) study

NSDUH has relied upon the availability of two individuals to provide the English and Spanish voices for the ACASI questions since 1999. If one of these individuals is no longer available when substantial changes are required, the entire survey may need to be rerecorded which may lead to an impact on trends. Text-to-speech (TTS) software could potentially provide a consistent voice for the ACASI portion of the survey without risk of interruption. The purpose of this study was to research the features and quality of various TTS software packages in order to determine if TTS software is an option for the NSDUH ACASI "voice." Based on a review of available software in the field, it appears that the time and expense necessary to create the customized pronunciations required for the NSDUH outweigh the potential benefits. At the moment, TTS software does not appear to be sophisticated enough for large-scale implementation, but this option may be revisited as the technology in this field progresses.

6. Questionnaire field test

The questionnaire field test, scheduled for 2011, is intended to examine whether and how the proposed questionnaire changes will affect the reporting of substance use behaviors, the overall timing of the questionnaire, and the amount of item-level nonresponse. It will also seek to determine whether interviewer and/or respondent debriefing questions provide proxy measures for data quality. The test will employ a 2x2 factorial design intended to yield 1875 interviews.

I. Other new studies related to the NSDUH redesign

1. Clinical validation study for dependence/abuse measures

This study will look into the measurement of substance abuse and dependence (SUD) for adults and adolescents through a clinical validation process once DSM-V criteria are finalized and published (expected 2013).

2. Data collection simplification and interviewer retention study

Currently NSDUH field staff are asked to perform a number of challenging activities. They have to persuade strangers to participate in an hour-long survey that contains questions about potentially illegal behaviors. This sometimes involves working odd hours, traveling to high-crime areas, dealing with combative respondents, and reassuring suspicious parents. This study presented data collection simplifications and protocol modifications that might increase interviewer morale and retention, and therefore reduce new-to-project interviewer training costs. Each idea was considered in terms of how the change might impact the survey should it be implemented, such as on the methodological integrity of the design, properties of the data, costs, and response rates. The study included field interviewer and field supervisor surveys focused on opinions about the current procedures and data collection simplification ideas. Several ideas from this study have been developed or incorporated into methods studies. Other ideas, such as the elimination of certain repetitive screener questions and increasing venues for field staff interaction, will be implemented without experimentation or extensive methodological testing.

J. Implementation issues

There are a number of issues beyond methodological concerns that many large scale-federal surveys must contend with when considering a redesign. A few of the major issues pertinent to NSDUH are discussed below.

Uncertainties and shifts in budget and agency priorities: One of the long established and collectively understood priorities of NSDUH has been the ability to measure trends in drug use and mental health problems annually. However, as the costs for collecting these data continue to grow, it is not always possible for the budget to continue to support these data. While the Office of Applied Studies (OAS) within which NSDUH is housed is a statistical unit under CIPSEA, OAS resides within a larger agency where the provision of services is considered more critical than the collection and dissemination of accurate data. This sometimes creates conflicting priorities among various managers, staff, Congress, and other constituents. It also increases the potential for unexpected shifts in agency priorities and budget for statistical activities, making it difficult to schedule and plan a redesign. As a result one of the goals of the redesign work is to be able to understand the consequences of such changes and to be able to design a survey where changes can be made without affecting trends.

Communicating technical issues to management: Survey practitioners face an additional challenge of communicating the technical impact of decisions to management in a way that they understand. This is especially important in situations where managers have limited expertise in statistical methods and survey research, but may be required to make formal presentations or have informal discussions with policymakers, constituents, and higher level managers about the pros and cons of redesign options. Concepts such as the effects of conducting an annual survey once every other year on estimates requiring pooled multi-year data; data collection being a primary driver of costs; the measurement and operational challenges associated with collecting biospecimen data; the nature of interviewer effects on key outcome measures; the need for advance planning time; the value of methodological work; and the challenges related to consistently measuring nebulous concepts and changing behaviors, have to be communicated appropriately to management to a point where it is salient to them, to ensure appropriate decisions are made.

Preserving trends between old and new design: A major concern of management and data users, and a primary goal of the redesign effort is the preservation of trends between the old and new design. One way of preserving trends is to create some sort of bridge between the old and new designs such that estimates from the old design can be adjusted to be comparable to the new one. In theory, this would involve a split sample utilizing both the old and new design including the changes in the sample design, questionnaire, incentives, contact materials, etc.

Attempts have been made in the past, with varying degrees of success, to maintain trendlines when implementing changes in methodology. For example, a methodology for adjusting past estimates to a new trend line was developed for the 1994 transition to a new questionnaire (Folsom, Jeffries, and Witt 1995). A random subsample was administered the old instrument and statistical adjustments were developed from analyses of the split-sample data. These adjustments were then applied to estimates from 1993 and earlier surveys.

In 1999, the survey moved from paper and pencil self-interviewing (PAPI) methodology to the new computer-assisted interviewing (CAI) methodology. A split sample approach was again used in an attempt to be able to separate changes in the key measures from the effects of the methodological changes and apply adjustments to the pre-1999 estimates for comparisons. However, at the same time, a new 50-state design was implemented and the combination of a much larger sample requirement along with the requirement for a bridge sample resulted in the recruitment of an inexperienced

interviewing force for a large part of the work. Because of the change in the mix of interviewer experience, the estimates of trend based on the paper and pencil methodology alone were suspect (Hughes, Chromy, Giacolletti, and Odom 2002). As a result, it was not possible to establish a bridge between the estimates before and after 1999 and a new baseline for trends was established starting in 1999.

In 2001, to improve response rates, incentives were considered for the first time. An experiment was designed to test the impact of monetary incentives ($0, $20, or $40) on response rates and prevalence rates. Based on analyses of the early results from the experiment, an incentive of $30 was offered to interview respondents beginning with the 2002 survey. At the same time, the name of the survey was also changed from the *National Household Survey on Drug Abuse* to the *National Survey on Drug Use and Health*; and certain other improvements in field protocol were developed and implemented during the course of the year (2002). Even though the 2001 experiment showed minimal effects of incentives on prevalence measures based on the relatively small sample (Wright, Bowman, Butler, and Eyerman 2002, OAS 2002), the analysis of the much larger 2002 national sample showed some unexpected and counterintuitive trends in prevalence measures. These changes in trends could not be directly associated with individual changes in procedures implemented in 2002, but the overall impact was considered large enough to justify starting a new trend line in 2002.

The NSDUH is a large-scale survey, and small differences in estimates can be both statistically and substantively significant. With such a large sample, even small changes in methodology can result in a difference that is large enough to affect interpretation of the data. In addition, research has suggested that sensitive measures collected in NSDUH are affected by survey conditions (e.g., interviewer experience, mode, setting, context, question wording) more than less sensitive items. Thus, survey designers have to be very careful when making methodological changes. One way to balance the need for trends with the need to improve how we conduct the survey is by attempting to isolate the effect of methodological changes.

With possible changes to the sample, questionnaire, data collection procedures, and estimation procedures, there is a strong chance that the redesign features will impact estimates of substance use and other behaviors. Some of the challenges to creating an effective cross walk between the new survey design and prior years include cost, the difficulty in implementing two different procedures in the field to collect the data and the possibility of cross-sample contamination. In particular, substantial budget cuts by the time of the redesign would require a major cut in sample size and a sample redesign, resulting in a change in field conditions. A new sample design can bring about changes to segment sizes, distance traveled by interviewers, the number of interviewers (and therefore field composition), hours worked, field supervisory structures, etc. These elements all have the potential to impact trends. Budget reductions will also limit the ability to field a split-sample.

Maintaining trend measurement after implementation: A goal of this redesign process is to develop a flexible survey design in which future changes to the sample design and questionnaire will not affect trends. In addition, the survey should be able to measure phenomena that are not static, such as changes in the uses of classes of drugs, slang terms to describe drugs, and the introduction of new drugs and cessation of use of other drugs.

Even if the decision is made to not attempt to isolate the effect of methodological changes on survey estimates, and instead establish a new baseline with the redesign, the challenge of measuring trends remains. The definition of ‘trend’ is an elusive one: Are trends truly being measured by using static methods of data collection and questionnaires, while the behaviors and phenomena being measured are continuously changing?

It is difficult to measure changes across time when the terms used to describe the same behaviors change; some drugs fall out of favor and new ones appear within classes of substances; new types of drugs outside the current classification system begin to be used; and usage patterns and drug combinations shift (e.g., marijuana and blunts). How can new variables be introduced in the survey without the inevitable spike that occurs with the introduction of new drugs in the questionnaire? How should those critical points of introduction for different drugs be determined?

Conventional wisdom may be that if we do not vary essential survey conditions (i.e., sample design, data collection methods, questionnaire, or estimation processes) then we should be able to maintain trends. But this assumption is not entirely valid. For example, a constant incentive may not have the same value over time to respondents. Inflation would decrease the value of the incentive, but on the other hand, a downward shift in the economy may lead to an upward movement in the value of the incentive. Shifts in public attitudes towards substance use, mental health, government, and surveys could also affect NSDUH trends.

One of the goals of this redesign effort is to design a survey that can withstand these elements. The plan is to build a framework of trend measurement that addresses changes in the behaviors and phenomena being measured as well as the practical issues faced by surveys such as budgetary constraints and shifts in priorities. Changes in the budget or data needs may lead to changes in the sample design, questionnaire or estimation methods. An additional goal is to develop a survey design that is flexible enough to incorporate improvements in areas such as data collection techniques, sampling and estimation.

As part of the redesign effort, one project is examining features in the current sample design that allow for flexibility, such as drawing additional primary sampling units (PSUs) to allow for supplemental studies or changing the sample design based on new estimation needs or resource availability. This project is searching for ways to incorporate an evaluation of the ability of these different designs to continue to maintain trends in the NSDUH, despite potential changes in field operations.

K. Lessons learned

Discussion and planning for the 2013 redesign of the NSDUH began in 2005. As of 2009, there is still a great deal more to do to reach the final determination of the new design, but much has been learned about the process that is applicable to other redesign efforts. Some of these lessons learned are discussed below.

Look to the past: The NSDUH project has over 30 years of institutional knowledge embedded within its data, documentation, and most importantly staff. For example, from past experiences we have learned that certain items are more susceptible to mode effects than others; that field tests due to small sample sizes do not always predict effects once the survey is scaled up; and that decisionmakers can unexpectedly alter the best laid plans at the last minute. All redesign preparatory work must keep these challenges in mind and, at the same time, attempt to predict what impact sample design or questionnaire changes will have on the data.

Look sideways: The NSDUH is not the only survey that has undergone a redesign. Several large-scale federal surveys such as the National Crime and Victimization Survey (NCVS), the National Household Education Survey (NHES), the Survey on Income and Program Participation (SIPP) and other surveys at the National Science Foundation (NSF), the National Center for Health Statistics (NCHS) and the Energy Information Administration (EIA) have redesigned or are looking to redesign their surveys. Much can be learned from the methods studies they have already conducted. In addition, many surveys currently use methods that the NSDUH redesign is considering and these surveys serve as a resource on both why those methods were adopted and how well they are performing. For example, as part of the redesign process we have examined how other surveys are using interviewer and respondent debriefing items as data quality measures; and the use of paradata in nonresponse weighting adjustments.

Look ahead: Given the lead time needed for implementing a redesign in a major survey such as NSDUH and the expectation that the new design will remain in place for many years, it is critical for the design team to try to anticipate future data needs and predict what problems are likely to evolve five or ten years in the future. One solution is a survey design that has flexibility, i.e., one that can transparently absorb changes in sample design and questionnaire content as budgets and data needs shift, and even as survey technologies change.

Consult with data users: One of the NSDUH expert panel members pointed out that no matter how well designed the new NSDUH is, it will not be accepted unless it can be presented as "new and improved". Part of this process would include the consultation with users about their questionnaire, estimation and data access needs. It also involves the identification of key stakeholders who will support the redesigned survey within their communities whether at the federal, state, academic or other level.

Obviously, consulting with data users can also result in a better designed, more relevant survey. However, input from data users must also be considered carefully. Often, suggestions for changes to the questionnaire do not reflect the needs of the greater community, but that particular data user's needs and research interests. In addition, not all data users understand the questionnaire design process and often suggest items that do not accurately measure the concept in which they are interested or are not possible to administer in the context of a large-scale household survey.

Keep management in the loop: Communicating plans and progress on the redesign clearly and regularly to senior management will help build their trust in the survey design team, allowing the redesign work to proceed more efficiently and effectively, without micromanagement by senior management.

Approach the redesign in its entirety; look for interconnections: When the redesign process first started, a number of distinct studies were proposed; each one isolated from the other. However, most of the parts of the survey are interconnected and must be considered together. Changing the sample design and sample size has effects on field staff, questionnaire (too small a subsample may make certain questions irrelevant due to small sample sizes), weighting, counting and listing, imputation procedures, etc. Examining weighting methods without considering the impact on field staff is not taking advantage of or maximizing the opportunity to redesign.

Keep key and secondary outcome measures in mind: While sometimes it is not easy to define what the primary and secondary (and possibly even tertiary) outcome measures should be and what precision level should be associated with them, early determination of these measures can then guide sample design and questionnaire content. The process of determining these key measures will involve management at different levels which has its own challenges in terms decision making.

Trust field tests; to a point: Field tests are small scale implementations. Often they use a select group of interviewers, have some cross-contamination with the main study, don't have adequate sample size, or might not be in all areas or languages covered by the main study. The lack of an effect between the original study and a redesign field test may not hold when the redesigned main study is implemented. This does not mean that field tests should not be conducted. It does mean that the link between analytic goals and field test design (e.g., what can we learn from our field test given sample sizes?) should be carefully examined prior to launching the field test.

Trust expert knowledge: Not everything is going to be tested—most surveys do not have the resources for the kinds of field tests and methodological studies that would be needed. In such cases, it is appropriate to trust expert knowledge, including experts in survey methodology who have experience in conducting large scale federal surveys in similar conditions, experts in the substantive areas of data collection who have experience with quantitative and measurement-related aspects, and experts by virtue of having a great deal of experience on the survey.

Think like a fed: As federal caretakers of these large, expensive and crucial data collections we have an obligation to seek counsel but evaluate all information through a lens that works best for the survey and data users. This is a unique perspective that only the federal survey managers can provide. No well-intended consultant or academic can ultimately understand all of the factors to be considered prior to making decisions regarding the design of a survey.

Anticipate failures: Despite the best intentions, careful planning and the most rigorous methodological work, respondents and data collectors do not always behave as we expect them to. Our goal during the redesign process is to minimize unanticipated problems. However, based on experience, it would be unrealistic to expect a redesign without unintended consequences that are not welcome.

L. References

Aldworth, J., Ault, K., Barnett-Walker, K., Carpenter, L., Copello, E., Frechtel, P., Licata, A., Liu, B., & Martin, P. (2007a, February). Imputation report. In *2005 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-0002, Deliverable No. 39, RTI/0209009.177.007). Research Triangle Park, NC: RTI International.

Butler, D. , Wright, D. , Chromy, J. , Meyer, M. , Granger, R. and Martin, P. (2003) *Effect of Incentives on Data Collection: A Record of Calls Analysis of the National Survey on Drug Use and Health* Paper presented at the annual meeting of the American Association for Public Opinion Research, Sheraton Music City, Nashville, TN

Gfroerer, J., Eyerman, J., and 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. (Web link: http://www.drugabusestatistics.samhsa.gov/redesigningNHSDA.pdf)

Grucza, R., Abbacchi, A., Przybeck, T., and Gfroerer, J. (2007), *Discrepancies in Estimates of Prevalence and Correlates of Substance Use and Disorders Between Two National Surveys* Addiction 2007;102(4):623–9. [PubMed: 17309538] (Web link: http://www3.interscience.wiley.com/cgi-bin/fulltext/117967863/PDFSTART)

Harrison, L. D., Martin, S. S., Enev, T., and Harrington, D. (2007). *Comparing drug testing and self-report of drug use among youths and young adults in the general population* (DHHS Publication No. SMA 07-4249, Methodology Series M-7).

Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies. (Web link: http://www.oas.samhsa.gov/validity/drugTest.pdf)

Hughes, A., Chromy, J., Giacoletti, K., & Odom, D. (2002). Impact of interviewer experience on respondent reports of substance use. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (DHHS Publication No. SMA 03–3768, pp. 161–184). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies. [Also available as a PDF at http://www.oas.samhsa.gov/redesigningNHSDA.pdf]

Hughes, A., Sathe, N., & Spagnola, K. (2009). *State Estimates of Substance Use from the 2006-2007 National Surveys on Drug Use and Health* (Office of Applied Studies, Substance Abuse and Mental Health Services Administration, NSDUH Series H-35, HHS Publication No. SMA 09-4362). Rockville, MD

Office of Applied Studies. (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.

Piper, L., Meyer, M., and Snodgrass, J. (2006 July) *2005 National Survey on Drug Use and Health: Reliability Study Pretest Final Report*. prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-0002, Deliverable No. 31, RTI/0209009.124). Research Triangle Park, NC: RTI International (Web link: http://www.oas.samhsa.gov/nsduh/2k5MRB/2k5relyTestFinal.pdf)

Ruppenkamp, J., Davis, T., Kroutil, L., & Aldworth, J. (2006). Methamphetamine analysis report. In *2005 National Survey on Drug Use and Health: Methodological resource book* (Section 20, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39, RTI 0208726.187.038). Research Triangle Park, NC: RTI International

SAMHSA (2007), *Review and Analysis of the National Survey on Drug Use and Health*. prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Task Order 280-03-1700. Westat, Rockville, MD.

Wright, D., Bowman, K., Butler, D., & Eyerman, J. (2002, May 16-19). *Nonresponse bias from the National Household Survey on Drug Abuse incentive experiment*. Presented at the annual meeting of the American Association for Public Opinion Research, St. Pete Beach, FL.