

Survey of Revenues and Expenditures (SRE) Supporting Statement

A1. Circumstances of Information Collection

The Substance Abuse and Mental Health Services Administration (SAMHSA) Center for Mental Health Services (CMHS) is requesting OMB approval to conduct the Survey of Revenues and Expenditures. This survey will collect comprehensive and reliable information from specialty providers of mental health (MH) and substance abuse (SA) treatment. The prevalence of mental illness and substance abuse disorders makes it vital to gain a better understanding of the size and character of our investment in treatment, identifying who is paying for services, and determining how much they are spending. The SSR&E is designed to fill the gap that exists since no existing studies provide direct revenue and expenses information at the facility level.

SAMHSA is specifically requesting OMB clearance of the SSR&E instrument, to be self-administered using paper or web (Attachment A), and an invitation to participate packet consisting of SAMHSA's invitation letter, Questions and Answers About the Survey, Thank you for Participating Overview (Attachment B). Note that instructions to facilities are incorporated into the questionnaire. The instrument (web and paper) will (1) note the authorizing legislation, (2) indicate the hours of burden expected, (3) will include the burden statement, and (4) display the OMB Number and expiration date.

Background and Legislative Requirements. Historically, SAMHSA has supported data collections used to develop an important tool that measures spending on mental health and substance abuse treatment and describes trends in financing care—the SAMHSA Spending Estimates (SSE) reports. The most recent SSE tracked spending from 1986 to 2003, uncovering and documenting previously unknown MH and SA provider and payer spending patterns, such as the rapid rise in Medicaid financing, the sharp decline in private insurance spending on SA treatment, and the explosive growth of MH prescription drug spending. In addition, the SAMHSA spending projections forecast a continued decline in the share of all health spending for MH and SA treatment. Current surveys, such as the National Survey of Substance Abuse Treatment Services (N-SSATS) and the National Survey of Mental Health Treatment Facilities (NSMHTF) either have low response rates or no longer collect financial information. This lack of current revenue and expense data is a critical gap in the information SAMHSA requires to achieve its goals.

SAMHSA last collected information on SA provider expenses and revenues through the N-SSATS more than a decade ago, in 1998. Similarly, SAMHSA collected information on specialty MH provider expenses and revenues through the Inventory of Mental Health Organizations (IMHO)/Survey of Mental Health Organizations (SMHO)¹ through 2004. These surveys no longer collect financial information; they are designed to collect non-financial data from each point-of-service location. Financial information was (and continues to be) often collected and maintained by a central office, thus, the point-of-service locations often did not have access to the requested information and/or did not have respondents who were knowledgeable about the

¹ IMHO/SHMO were later redesigned as the National Survey of Mental Health Treatment Facilities (NSMHTF). NSMHTF is currently being redesigned as the National Mental Health Services Survey – N-MHSS.

facilities' financial records. As a result, these surveys often yielded poor item response with inaccurate financial information. The absence of financial data creates a significant gap in information needed for the accurate estimation of MH/SA treatment spending. The gap currently is filled with imputed spending based on client counts and extrapolated or imputed cost-per-client data; however, that results in trend estimates for some providers that are inconsistent with information from other sources.

Two sections of the Public Health Service (PHS) Act authorize the SSR&E data collection: Section 505 (42 USC 290aa-4) and Title 42, Chapter 6A, Subchapter III-A Substance Abuse and Mental Health Services Administration, Part A Organization and General Authorities.

Overview of the Study Design

Two primary research questions guide this study:

1. How much is the nation spending on MH and SA treatment in specialty care facilities?
2. What are the expenses per client for treatment in MH and SA specialty care facilities?

The SSR&E survey, for which clearance is requested, is designed to explore these two primary research questions. The survey has four modules: (1) the types of services provided by a facility (for example, whether the facility provides inpatient and/or outpatient care), (2) the facility's sources of revenue (for example, how much of the total net revenue comes from Medicaid or client private insurance), (3) client counts (such as number of clients in inpatient services), and (4) facility expenses (such as total operating expenses). The final pilot tested version of the SSR&E is attached as Attachment C. SAMHSA pilot tested the instrument during April - September of 2009 with 9 facilities and made numerous iterative changes to it.

In order to examine these questions, SAMHSA will draw a sample of 2,000 facilities from existing frame data from the inventories of SA treatment facilities (N-SSATS) and MH treatment facilities (Inventory of Mental Health Organizations and General Hospital Mental Health Services [IMHO/GHMHS]). It is anticipated that approximately 1,500 facilities will participate.

A2. Purpose and Use of Information

Purpose of the Information Collection. SAMHSA expects a one-time data collection in 2010. SAMHSA and other policy makers will use the information to gauge the financial health of providers and the vitality of the industry. Financial statistics increase understanding of insurance and cost barriers that clients may face. They also assist SAMHSA in allocating resources to providers and developing strategies for improving public and private payer coverage. Finally, researchers and policy analysts will use this information to develop methods and public use files to project costs from routine facility surveys for some years to come.

Use of the Information. The primary use of the data collected by the SSR&E is to provide accurate updated estimates for future editions of SAMHSA's National Expenditures for Mental Health Services and Substance Abuse Treatment report. This report, which has been issued periodically since 1999, is regarded as the leading authoritative source for information about how much the United States has invested in MH/SA treatment (Mark 2007). The SSR&E also will provide information to support SSE trend analyses.

Finally, the new information collected by the SSR&E will also be used to improve survey results by providing ranges for validating survey responses through the analysis of client case load, case mix, and staffing by mode of care for peer-group facilities. These data can improve imputation techniques by furnishing information on facility type; mode of care (inpatient, outpatient, residential); payment source; and diagnosis (case mix) that has proven useful in the past for imputing values for item and facility nonresponse. The data can also meet other important SAMHSA informational needs, including a more thorough understanding of factors such as:

- Expenses and revenue sources of different types of facilities with different diagnostic case mixes or treatment capabilities (MH or SA, or MH and SA), treatment settings (inpatient, outpatient, and/or residential), and staffing models.
- The total cost of treatment, labor, and other operating expenses as an adjunct to cost bands currently used by SAMHSA to evaluate discretionary grant awards aimed at improving quality of care and client retention, or targeting care to specific vulnerable client populations (for example, linking at the facility level with the N-SSATS and N-MHSS, the SSR&E can be used to study populations with HIV, both MH and SA disorders, pregnancy, or alcoholism).
- The results of the SSR&E survey will also support the estimation of revenues and expenditures for facilities that are part of SAMHSA's N-SSATS and N-MHSS surveys for several years into the future, using imputation and projection techniques.

Data on revenues and expenses have not been collected from SA or MH facilities for a decade. This hiatus has left a significant gap in SAMHSA's knowledge of revenues and expenses.

A3. Use of Information Technology

To lessen respondent burden, participants may respond through the web instrument, on paper (mail or fax), or through telephone assistance with the contractor's facility liaison. The web instrument will offer the easiest means of providing data as it will be programmed to calculate sums and to display definitions of terms and lists of state S-CHIP and Medicaid names. This will allow the facility to respond without resorting to searching through the paper instrument for definitions or resorting to use of their own calculators. The instrument will improve data quality by automatically skipping to the next appropriate question based on responses. Based on other establishment studies, we expect that at least 25 percent of facilities will choose to use the web version of the instrument.

SAMHSA will implement the OMB survey instrument using the contractor's WebSurv.NET software.

A4. Efforts to Identify Duplication

Although no studies have collected similar data for close to a decade, SAMHSA recognizes certain questions asked on prior surveys of revenue and expenses data remain relevant, valid, and reliable for this study's primary purpose. Where possible, SAMHSA has taken questions from other earlier SAMHSA facility surveys such as the National Survey of Substance Abuse Treatment Services (N-SSATS), National Survey of Mental Health Treatment Facilities (N-SMHTF), the Alcohol and Drug Services Study (ADSS), and the Uniform Facility Data Set (UFDS), and the Inventory of Mental Health Organizations (IMHO) when the surveys asked for revenues and expenses data. Attachment D (SSR&E Question Sources) lists each question the SSR&E asks and indicates its source and reason for inclusion. In no instance are questions duplicated across current SAMHSA surveys, with the exception of those items needed to insure which organizations and facilities can be identified across our surveys.

A5. Involvement of Small Entities

The SSR&E will collect data from MH/SA treatment facilities that vary greatly in size from small independent service delivery operations to units within large hospitals. To minimize burden on small entities, the questionnaire will be available through a paper and a web version and facilities may choose how to access it. We expect that the facility staff best suited to respond to the survey will be people in the financial area who track costs, such as the Chief Financial Officer. Very small businesses may not have a person in this specific role but out-source their accounting services. Contractor staff will be trained to assist small facilities in providing information, for example, by explaining to relevant staff exactly what information is required to answer each question or by speaking with relevant individuals. Contractor staff will also be trained to help small businesses determine how to fit their financial information into the categories used in the survey. In order to minimize burden on small facilities, the survey attempts to request data in a manner that reflects how it is collected by the facilities. Investigating the issue of how financial information is collected and stored by the facilities was an important part of the pilot test.

A6. Consequence If Information Collected Less Frequently

If data are not collected, SAMHSA will have no direct information about MH/SA facilities revenues and expenses. Important SAMHSA reports such as the SAMHSA National Expenditures for Mental Health Services and Substance Abuse Treatment report will continue to lack accurate financial information on MH/SA specialty facilities and will have to rely on imputations based on ten-year-old data. SAMHSA expects to field the SSR&E only once.

A7. Consistency With the Guidelines in 5 CFR 1320.5(d)(2)

This data collection is consistent with the guidelines in 5 CFR 1320.5(d)(2).

A8. Consultation Outside the Agency

a. Federal Register Announcement

The 60-day FRN required in 5 CFR 1320.8(d) was published in the *Federal Register* on July 13, 2009 (Vol. 74, page 33446.). On September 10, 2009, SAMHSA received public comments and responded to them on October 2, 2009. The comments and SAMHSA's responses are found in Attachment E. The comments suggested a number of changes, tightened the use of terminology in the questionnaire, clarified funding categories and made them consistent across SSR&E and N-MHSS, clarified whether justice system funding was from juvenile or criminal justice, and reinforced our decision to omit the section on staffing.

b. Consultation Outside SAMHSA

SAMHSA convened a Technical Expert Panel (TEP) attended by the following technical experts who represented MH/SA treatment facilities, hospitals, health associations, and staff from other federal agencies. (See Attachment F for details about the TEP members.)

As a result of the consultation, the survey has been clearly focused on finances and has omitted a number of questions the TEP felt would be burdensome to answer.

A9. Payment to Respondents

No payment or gift will be provided to respondents.

A10. Assurance of Confidentiality

No information will be collected about identifiable individuals. The data will be collected about the characteristics of a treatment facility. No assurance of data privacy will be pledged. Because this survey does not involve human subjects, Institutional Review Board (IRB) clearance will not be sought. All data will be maintained in a password protected data system.

A11. Questions of a Sensitive Nature

SAMHSA is not collecting information of a sensitive nature from facilities.

A12. Estimates of Annualized Hour Burden

After pilot testing, SAMHSA made revisions which substantially reduced respondent burden. SAMHSA bases the three – hour annualized burden on the results of the pilot test, in

which respondents (one or more) required a maximum of 3 hours to complete. SAMHSA will collect data from financial experts at 1,500 facilities, amounting to a total of 4,500 hours. Given the range in size and operating budgets of facilities, SAMHSA calculated the average hourly rate of respondents by averaging the mean hourly rate of financial managers and the mean hourly rate of accountants and auditors (according to the Bureau of Labor Statistics' industry-specific occupational employment and wage estimates for outpatient mental health and substance abuse centers). Assuming that these occupations have a combined average hourly rate of \$33.03, the total cost to all facilities responding to the survey will be \$148,635. SAMHSA is conducting a pilot test to clarify the exact hours needed to complete the SSR&E instrument.

TABLE 1

BURDEN

Form	Number of Respondents	Responses Per Respondent	Total Responses	Hours per Response	Total Hour Burden	Hourly Wage Cost	Total Hour Cost
SSR&E	1,500	1	1,500	3	4,500	\$33.03	\$148,635

A13. Estimates of Annualized Cost Burden to Respondents

There are no startup, operational, or maintenance costs for the respondents associated with this data collection.

A14. Estimates of Annualized Cost to the Government

The annualized cost to the government is calculated to be \$634,103 per year.

A15. Changes in Burden

This is a new data collection.

A16. Time Schedule, Publication and Analysis Plans

a. Time Schedule

Over a six-month data collection period, SAMHSA expects to attain a 75 percent response rate: 25 percent completed by web, 30 percent completed by paper (mail or fax), and 20 percent completed through telephone assistance with the facility liaison.

The survey design involves several important scheduled steps: (1) encouraging facility participation through an early marketing campaign, (2) training data collection liaison staff to facilitate facility participation, (3) coordinating advance mailings and calls to identify best

respondents (usually financial officers) and further encourage participation, (4) conducting follow-up calls to encourage and assist facilities in providing information and to retrieve missing information, and (5) implementing the OMB-approved SSR&E questionnaire in both a paper format and a web format. The data collection timeline is presented below in Table 1. The contractor will use interviewing staff who are trained on both the N-SSATS and NSMHTF surveys, are highly experienced working with both MH and SA facilities, and have a clear understanding of the services that are provided. The schedule will change depending on when OMB clearance happens.

TABLE 2
PLANNED SURVEY SCHEDULE

Activity	Start Date	End Date
Marketing campaign	01/06/09	02/05/10
Implement instrument in WebSurv	1/08/10	2/15/10
Train for advance calls	2/15/10	2/16/10
Train for follow-up calls	3/15/10	03/16/10
Begin data collection	2/17/09	8/17/09
Coordinate mailing and calls	2/17/09	04/23/10
Conduct follow-up calls	4/23/10	08/17/10
Maintain tracking system	2/17/10	8/17/10
Clean and prepare analytic files	06/01/10	08/26/10

b. Tabulations

Basic descriptive tabulations on all survey data elements will be produced, including sums, means, medians, distributions, and standard errors for each data element, as appropriate. A small number of standard descriptive categories will be defined that may include:

- Type of facility (for example, multiservice mental health facility, specialty or general hospital).
- Diagnoses treated (MH, SA, or both).
- Ownership/profit status (public federal, public state/local, private not-for-profit, or private for-profit).
- Services provided (for example, inpatient only, residential only, outpatient only, multi-mode).

These analyses will provide an overview of the survey results and serve as a statistical reference for analysts and policymakers.

SAMHSA and its contractors will also identify areas on which to develop focused studies such as:

- The primary study will be an assessment of the survey data as a source for SAMHSA's national spending estimates for mental health and substance abuse treatment. Revenues will be analyzed by payment source and setting of care for national estimates based on this survey and others data sources.
- Another focus will analyze expense per client served to understand the costs of treatment for clients with mental and/or substance use disorders in different settings of care and types of facility. Expenses per admission can also be analyzed by linking the SSR&E to the larger facilities which collect admission and client counts.
- More generally, because the SSR&E can link to the other facility responses, many other topics can be explored using the content of the N-MHSS and N-SSATS facility surveys and the overall financial data from the SSR&E.

c. Publications

SAMHSA will publish reports for the focused studies, including discussion of importance of the issue from a national policy perspective, survey and analytic methods, results, conclusions, and related policy implications.

Preliminary Report of Weighting Approach and Revenue Results. SAMHSA plans to produce a 50-page preliminary report describing methods for extrapolating survey findings to the universe of MH/SA facilities. This task will include developing a method for using data from the SSR&E sample to extrapolate to the two national census surveys on mental health and substance abuse facilities being fielded separately. This weighting will support the estimates by the type of provider, type of care, and source of payment categories used in SAMHSA's national spending estimates for MH/SA treatment. The preliminary report will consist of five analyses: (1) The derivation of facility weights for national estimates, (2) analysis of single mode of care providers (i.e., those that provide only one type of treatment (inpatient, outpatient, or residential)) for MH, SA, and combined MH/SA treatment; (3) analysis of multi-mode of care providers; (4) imputation of payment source distribution when necessary; and (5) aggregation across facilities by diagnosis, type of provider, type of care, and payment source.

Final Report of Survey Findings, Final Data File, and Data Dictionary and Documentation. SAMHSA contractors will prepare a final report with clear, descriptive information that will enable technical and nontechnical audiences to quickly and easily access the information relevant to their area of interest. We expect a varied audience for the report, including state administrators of MH/SA services, members of organizations representing providers of specialty behavioral health services, individual health services providers, MH/SA researchers, and MH/SA advocacy organizations.

The final report will contain: (1) executive summary of findings; (2) background of the survey; (3) survey methodology; (4) findings; (5) implications of the findings; (6) recommendations for further research; and (7) appendices with names of TEP members, some technical tables, and the survey itself. The executive summary will present a clear and concise summary of the full report and highlight key findings and recommendations. The background section will describe the context of this project, emphasizing the rationale and use for data on national revenues and expenses of specialty MH/SA treatment providers and the absence of a regular, reliable, comprehensive collection of these data. This introductory section also will describe the survey's purpose and research questions of the project. The survey methodology chapter will describe the characteristics of the sample and discuss the development of the questionnaire and the data-collection processes, including the advance contact, interview, and follow-up procedures. This section will include a description of the imputation and weighting methods. It also will provide the survey results, such as response rates tabulated by type of diagnosis treated at the facility (MH, SA, or both); type of services provided (inpatient only, residential only, outpatient only, and multi-mode); type of facility (general hospital, specialty hospital, or other specialty MH/SA facility); and type of ownership (federal, state/local, private not-for-profit, private-for-profit). The findings section will synthesize the data and describe the overall findings. The findings will address the research questions. Data tables, with the survey data tabulated by type of diagnosis, type of service provided, type of facility, and type of ownership will be included as statistical reference. These basic descriptive tabulations will include sums, means, medians, distributions, and standard error as appropriate. We will discuss in a separate chapter the implications for policymakers aiming to allocate federal support for treatment services based on the financial status of these facilities. The last section of the report will explore new areas for research or areas for more detailed research, such as staffing, wage rates, and efficiency of facilities.

In addition to the reports, SAMHSA contractors will prepare a de-identified public use file (PUF) and documentation. The PUF data will be housed at SAMHSA. The PUF will be accompanied by documentation containing detailed information for each data variable and summary information for the survey as a whole. Data will be provided in ASCII format, with example programs for loading the data into the SAS, SPSS, and STATA statistical software packages. The documentation will be provided as machine-readable Adobe PDF files and will be 508 compliant.

The final report and data will be available in approximately April 2011.

A17. Display of Expiration Date

SAMHSA will display the OMB number and expiration date on both the paper and web versions of the questionnaire.

A18. Exception to the Certification Statement

The certifications are included in this submission.

B1. Respondent Universe and Sampling Methods

a. Respondent Universe

The primary goal of the sampling design is to support the collection of detailed revenue and expense information from a national random sample of facilities that offer mental health (MH) treatment services, substance abuse (SA) treatment services, or both. The universe for this study is essentially all treatment facilities in the United States that provide these treatment services.² At this time, there is no single source list for this population. However, separate surveys (actually censuses) currently are being conducted of all MH treatment facilities and all SA treatment facilities.³ The two surveys are the National Survey of Mental Health Treatment Facilities (NSMHTF) and the National Survey of Substance Abuse Treatment Services (N-SSATS). Because some facilities provide both MH treatment services and SA treatment services, the samples and sampling frames for these two surveys overlap. For the SAMHSA Survey of Revenues and Expenses (SSR&E), constructing the sampling frame will involve merging the frames for the N-SSATS and the NSMHTF, parsing out the unique facilities, and identifying facilities that provide only MH treatment services, only SA treatment services, or both.

A description of the two surveys follows.

3. **National Survey of Substance Abuse Treatment Services (N-SSATS).** The N-SSATS is an annual survey which includes all facilities that offer SA treatment services (currently numbering 16,284 facilities). The frame for the N-SSATS was developed from the Inventory of Substance Abuse Treatment Services database, or ISATS. The ISATS is a listing of all known public and private SA treatment and prevention facilities in the U.S. and its territories. The database for ISATS is maintained by Synectics for Management Decisions, Inc. under contract with the Office of Applied Studies (OAS). The ISATS is updated monthly to account for the changing universe of facilities offering SA treatment services. The N-SSATS frame is a subset of the ISATS database; it excludes facilities from the ISATS that (1) are closed, (2) provide only prevention services, (3) provide only administrative services, (4) are jails/detention centers, (5) are halfway-house-only facilities, or (6) are solo practices that are not approved by their state to be in the N-SSATS.⁴ For the SSR&E, these general hospitals that do not have specialized SA units and all solo practices (even if they are state-approved to be in the N-SSATS frame) are not part of the target population and will be excluded during the sampling frame construction or filtered out in the SSR&E questionnaire.

4. **National Survey of Mental Health Treatment Facilities, or NSMHTF.** The NSMHTF is the redesign of a biennial survey, called the Survey of Mental Health

² Excluded from this universe are halfway-house-only facilities (that is, facilities that do not provide treatment services at their location), jails/detention centers, general hospitals that do not have specialized SA and/or MH units, and solo practices.

³ These surveys do not collect the detailed data needed for this study.

⁴ Solo practices in N-SSATS are treated differently than other practices. The Substance Abuse Agency of each state decides what list of solo practice substance abuse treatment facilities in that state should be included in the N-SSATS. For some states, this list is limited to licensed facilities; for other states, it includes all facilities that provide SA treatment services. Some facilities that are supposed to be excluded inadvertently end up in the N-SSATS frame and are filtered out in the questionnaire. They are excluded from the frame in the next survey year.

Organizations (SMHO), which collected data for mental health organizations. For 2008, NSMHTF was redesigned to make the individual facility the unit of analysis. The NSMHTF includes all facilities that have units specifically geared toward MH treatment services, (currently 13,201 facilities)⁵ and excludes all solo practices, as well as facilities that are general hospitals with scattered beds dedicated to MH, but no specialty unit for mental health services. The sampling frame for the NSMHTF is called the Mental Health locator file. This locator file was developed using an inventory of MH treatment facilities called the Inventory of Mental Health Organizations and General Hospital Mental Health Services (IMHO/GHMHS). The IMHO/GHMHS is a biennial, complete enumeration of all specialty MH organizations and separate psychiatric services of non-federal general hospitals in the U.S. For the 2008 NSMHTF, this inventory was maintained by Social and Scientific Systems, Inc. under contract with SAMHSA's Center for Mental Health Services (CMHS). In 2010, the NSMHTF will be renamed the National Mental Health Services Survey (N-MHSS) and will be conducted by MPR.

The sampling frames for these two surveys will be combined into a single frame that includes both MH and SA treatment facilities. The two frame files have their own unique facility level identification numbers, and new identifiers will be developed for each facility in the merged file. To account for the overlap in the two sampling frame sources, several methods will be used to identify facilities in both files.

b. Sampling Methods

The purpose of the SSR&E is to collect detailed revenue and expense information from a national random sample of treatment facilities to support development of reliable revenue and expense estimates for all treatment facilities. For the validity of these estimates, it is important to ensure representation on a series of factors. Among these factors are (1) mode of care, (2) type of facility, and (3) facility size. These three factors will be used for sample stratification. We will select a stratified random sample of 2,000 facilities from the universe of facilities described in Section B1a to obtain data from 1,500 facilities (assuming a 75 percent completion rate).

(1) Explicit Stratification Variables

As described above, the following variables will be used for explicit stratification:

- Mode of care
- Type of facility
- Facility size

Strata will be created by cross-classifying the levels of these variables, collapsing levels where necessary. The anticipated strata created by cross-classifying these variables will be described in Table 4, in Section B1.b (3). First, however, we describe each of these variables and their levels in turn, how we define stratum boundaries for facility size, and the anticipated client counts in each stratum.

Mode of care is considered the most important of the variables, and will be available for most facilities from the sampling frames. Four categories are anticipated for this stratification variable: (1) outpatient-only facilities, (2) residential-only facilities, (3) outpatient and residential facilities, and (4) facilities that offer inpatient care either alone or in addition to other modes of

⁵ This figure excludes 492 administrative facilities.

care. Preliminary counts of different modes of care by SA treatment facilities (from the 2006 N-SSATS) and by MH treatment facilities (from the 2008 NSMHTF locator file) are shown in Table 1. This table shows that there are a relatively large number of facilities in the outpatient-only, residential-only, and outpatient plus residential categories. However, there are relatively few facilities offering inpatient care among SA treatment facilities or MH treatment facilities, either as the only mode or in combination with other modes of care. Because of this, all of the mode of care categories with inpatient treatment services will be collapsed into one stratum. The category listed as “General Mental Health Services” in Table 1 refers to facilities where the mode of care is unknown.⁶

It should also be noted that the definition of “residential” varies between the SA and MH communities. Because mode of care is available on both frames for most facilities, it will be possible to ascertain the degree to which the definition of “residential” differs between the two frames. If a large difference in definition is evident, it may be necessary to separate, where feasible, SA facilities and MH facilities offering residential care into different strata.

The other two stratification variables—facility type and facility size—are not available on the sampling frames. These variables will have to be obtained directly from the 2008 NSMHTF and the 2008 N-SSATS surveys. The accuracy and completeness of these data are dependent on the response to these surveys. The response rate for the 2008 N-SSATS is very high (exceeding 94 percent), so even with some item nonresponse, we expect to obtain facility size and facility type information for most SA facilities. However, the response rate for the 2008 NSMHTF will be substantially lower (perhaps as low as 60 percent). Hence, facility type and facility size will not be available for a large portion of facilities that offer only MH services, resulting in strata where one or both of these variables is “unknown.” Because the combined frame has not yet been developed, precise information on the size of strata in the respondent universe defined by these variables is not yet available. However, it is likely that cross-classifications of these variables will require some collapsing of levels when defining the strata.

⁶ We expect that some of the facilities currently categorized under General Mental Health Services will not respond to the survey. If the number of facilities with unknown mode of care is large enough, a separate stratum may be required with unknown mode of care.

TABLE1. MODE OF CARE AMONG SELF-REPORTING FACILITIES OFFERING SA AND MH SERVICES

Mode of Care ^a	Number of SA Facilities ^b	Number of MH Facilities ^c
Outpatient only care	8,501	6,433
Residential only care	2,323	1,784
Outpatient + residential care	1,066	824
Inpatient only care	284	959
Inpatient + residential care	37	53
Inpatient + outpatient care	346	875
Inpatient + outpatient + residential care	134	150
One or more modes unknown	26	N/A
General mental health services	N/A	2,123 ^d

Source: ^aIn the 2006 N-SSATS, there were 1,054 SA facilities for which the inpatient, outpatient, and residential client counts were reported by other facilities and 661 where the reports included multiple facilities. For the purposes of this table, we exclude the 1,054 unreported facilities, and we include the data for the 661 which included information for multiple facilities. For the SSR&E, each of the individual treatment facilities will be in the sampling frame, regardless of how their attributes are reported in the N-SSATS.

^b2006 National Survey of Substance Abuse and Treatment Services.

^c2008 National Survey of Mental Health Treatment Facilities locator file.

^dMental health treatment facilities for which mode of care is unknown at this time.

For the type of facility, four categories are anticipated: (1) general and Veteran’s Administration (VA) hospitals with specialty units, (2) specialty psychiatric and SA hospitals, (3) specialty community MH facilities, and (4) specialty community SA facilities. We will use the primary focus of care to differentiate between specialty community MH facilities and specialty community SA facilities. However, if the primary focus of care is a mix of both MH and SA treatment services, a separate stratum for specialty community MH/SA facilities may be developed. Primary focus of care will not always be directly available, though it can be ascertained for most facilities.

Facility size is highly correlated with important outcome variables involving revenues and expenses. Because of this, it is likely that the values for these variables will substantially differ in subpopulations defined by facility size; therefore, it will be used as an explicit stratification variable. The number of categories defined by this variable will vary depending upon the level of mode of care and facility type. Two variables will be used to measure facility size: number of beds and client counts, with one or both variables used, depending upon the mode of care and facility type. These variables are not available on the sample frames, though they are available on the N-SSATS and NSMHTF surveys.

The client count variable will be used to measure size for most strata, including strata involving all types of hospitals, and for strata involving community MH and SA treatment facilities that either (1) offer inpatient care or (2) offer only outpatient care. It will also be used in conjunction with the number-of-beds variable for non-hospital residential treatment facilities.

The client count variables in the two surveys are very similar. Both surveys ask about the number of clients that received MH (for the NSMHTF) or SA (for N-SSATS) services on a single day for each mode of care (inpatient, residential, or outpatient). For the NSMHTF, that day is April 30, 2008; for N-SSATS, that day is March 31, 2008.

The number-of-beds variable will be used for strata involving non-hospital residential treatment facilities that do not offer inpatient care, with a breakpoint of 16 or fewer beds versus 17 or more beds. This breakpoint is used because current federal law prohibits Medicaid reimbursement for any person over age 21 and under age 65 who resides in an institution for mental diseases (IMD), even for treatment unrelated to mental illness. An IMD refers to a hospital, nursing facility, or other institution of more than 16 beds that is primarily engaged in providing diagnosis, treatment, or care of persons with mental diseases. This is commonly referred to as the “IMD exclusion.” The IMD exclusion applies to most general and specialty hospitals because hospitals with fewer than 17 beds are very rare. That is why this stratification is useful only for size stratification of non-hospital residential treatment facilities that do not offer inpatient care. If a non-hospital residential treatment facility stratum with 17 or more beds ends up being quite large, we will use the client counts variable to further divide the stratum.

In the N-SSATS and NSMHTF surveys, the number-of-beds questions are slightly different. In the 2008 NSMHTF, the question asks about the number of beds set up and staffed for the treatment of mental illness, whereas in the 2008 N-SSATS, the question asks about the number of beds specifically designated for substance abuse treatment. The N-SSATS definition of “number of beds” is much narrower than the NSMHTF definition; it appears that the IMD exclusion most closely fits the NSMHTF definition, so this is the definition we will use.⁷ When the data for both surveys are available, it will be possible to evaluate the relationship between the two number-of-bed variables that are derived from the two surveys. It may be necessary to adjust the results from the N-SSATS bed counts to more closely match what the NSMHTF bed count is measuring.

(2) Explicit Stratum Boundaries for Continuous Variables

Boundaries for strata based on facility size, a continuous variable, will be determined using the cumulative square root rule (Cochran, 1977) that was developed under the assumption of a Neyman allocation. The Neyman allocation is a method of allocating sample to strata by allocating a proportionally larger sample to strata with a lot of variability, and relatively less sample to strata with less variability. Preliminary calculations using data from N-SSATS within modes of care indicate that large facilities will have to be selected with certainty to reduce the variance. The other non-certainty strata will be constructed using this algorithm to improve the efficiency of the sample.

(3) Proposed Strata: Population and Sample Counts

Because data for the MH treatment facilities are not yet available, it is not possible to make a definitive choice of strata. However, based on the data that are available, it was possible to estimate the distribution of the stratification variables. For example, data from the 2006 N-

⁷ For example, using the N-SSATS definition, if a facility offers multiple services but does not designate beds specifically for SA services, it is conceivable for a fully functional SA treatment facility to have 0 beds specifically designated for SA treatment.

SSATS and the OMB package for the 2008 NSMHTF were combined to obtain an estimate of the total number of facilities (25,380) classified by facility type in Table 2. Data from the 2006 N-SSATS were used to determine the distribution of facility type within each mode of care for SA facilities in Table 3. We used this information to anticipate the strata shown in Table 4, based upon the assumption that there are 25,490 total facilities (slightly more than the total assumed in Table 2).

TABLE 2. DISTRIBUTION OF MENTAL HEALTH AND SUBSTANCE ABUSE TREATMENT FACILITIES AND HALF-WIDTHS OF 95 PERCENT CONFIDENCE INTERVALS FOR ESTIMATES OF TOTAL REVENUES IN SELECTED SUBPOPULATIONS, PRESENTED AS PERCENTAGE OF ESTIMATE

Subpopulation of Interest (Type of Facility)	Number of Facilities in U.S. ^a	Proportion of all Facilities	Expected Number of Facilities in Sample	Half -Width CI Percent of Mean
General and veteran's affairs hospitals with specialty MH/SA units	2,050	0.081	120	5.37%
Psychiatric and substance abuse specialty hospitals	1,130	0.045	66	7.24%
Specialty community mental health facilities ^b	11,000	0.433	651	2.30%
Specialty community substance abuse facilities ^b	11,200	0.441	663	2.28%
Total mental health and substance abuse facilities	25,380	1.000	1,500	1.52%

Source: ^a2008 OMB Package for National Survey of Mental Health Treatment Facilities and 2006 National Survey of Substance Abuse and Treatment Services

^bThese strata will be formed by identifying the primary focus of the community facilities as mental health, substance abuse, and both mental health and substance abuse. The latter would be facilities where neither condition is considered the primary focus because patients with principal diagnoses of either type are recruited or accepted and treated. This latter group would be a third category of community facilities. We do not have the data currently to allocate the sample accordingly, although the final stratification may contain such a category.

TABLE 3. NUMBER OF FACILITIES OFFERING SA SERVICES BY MODE OF CARE AND FACILITY TYPE^A

Mode of Care ^b	Facility Type			
	General and VA Hospitals	Psychiatric and Specialty Hospitals	Clinics and Health Centers	Unknown
Inpatient and combinations	487	204	109	1
Outpatient only	692	104	7,702	3
Residential only	51	18	2,254	0
Outpatient and residential	83	7	976	0
Unknown	5	1	20	0

Source: ^a2006 National Survey of Substance Abuse and Treatment Services

^bIn the 2006 N-SSATS, there were 1,054 SA facilities for which the inpatient, outpatient, and residential client counts were reported by other facilities and 661 where the reports included multiple facilities. For the purposes of this table, we exclude the 1,054 unreported facilities, and we include the data for the 661 which included information for multiple facilities. For the SSR&E, each of the individual treatment facilities will be in the sampling frame, regardless of how their attributes are reported in the N-SSATS.

Table 4. Proposed Strata: Population and Sample Counts

Stratum Number	Mode of Care	Facility Type	Primary Focus	Size strata	Estimated Population	Estimated Sample ^a
1-3	Inpatient	General & VA hospital with specialty unit	MH and/or SA	3 client count strata	250 per size strata	15 per size strata
4-6	Inpatient	Psychiatric & specialty hospitals	MH and/or SA	3 client count strata	250 per size strata	15 per size strata
7	Inpatient	Specialty MH facilities	MH & MH/SA	1 stratum	100	6
8	Inpatient	Specialty SA facilities	SA & MH/SA	1 stratum	100	6
9-11	Outpatient only	General & VA hospital with specialty unit	MH and/or SA	3 client count strata	350 per size strata	21 per size strata
12-13	Outpatient only	Psychiatric & specialty hospitals	MH and/or SA	2 client count strata	180 per size strata	11 per size strata
14-23	Outpatient only	Specialty MH facilities	MH & MH/SA	10 client count strata	775 per size strata	46 per size strata
24-33	Outpatient only	Specialty SA facilities	SA & MH/SA	10 client count strata	775 per size strata	46 per size strata
34	Residential	All hospitals	MH only	1 stratum	100	6
35	Residential	All hospitals	SA only	1 stratum	100	6
36	Residential	All hospitals	MH and SA	1 stratum	100	6
37-40	Residential only	Specialty MH facilities	MH & MH/SA	4 bed/client count strata	550 per size strata	32 per size strata
41-44	Residential only	Specialty SA facilities	SA & MH/SA	4 bed/client count strata	550 per size strata	32 per size strata
45-48	Outpatient/residential	Specialty MH facilities	MH & MH/SA	4 bed/client count strata	250 per size strata	15 per size strata
49-52	Outpatient/residential	Specialty SA facilities	SA & MH/SA	4 bed/client count strata	250 per size strata	15 per size strata

^aThese figures assume proportional allocation, which will be sufficient for determining the allocation within strata defined by mode of care and facility type. However, the number within each of the size strata will not be equal (as shown here). Instead, the number in each of the size strata will vary according to a Neyman allocation, as explained in Section b.2.

A sample of size 1,500 was selected because this was the smallest sample size possible to obtain the necessary precision for subpopulations of interest. In addition, there would be sufficient data available for the purposes of combining revenue and expense data with data from other surveys using regression models (especially for subpopulations that are too small to obtain useful estimates).

Most subpopulations of interest are in separate strata. The exceptions: facilities offering only residential care and those offering residential plus outpatient care are collapsed in strata 34-36, as are general hospitals and specialty hospitals. After the sampling frames are merged, we will review this stratification and determine if changes are appropriate to ensure the desired counts are obtained in key subpopulations.

(4) Implicit Stratification

A sequential sampling scheme will be used to include implicit stratification variables as well as explicit stratification variables. Possible implicit stratification variables include geographic location, ownership of facility, primary focus of care (where not used for explicit stratification), and specific facility size (within facility size categories).

(5) Allocation of Number of Facilities per Explicit Stratum

Proportional allocation was used to assign the number of facilities to be selected within each stratum in Table 2.⁸ Proportional allocation will be sufficient for explicit strata defined by mode of care and facility type. However, the Neyman allocation, which accounts for differing variances across strata, is “usually superior to proportional allocation in populations in which gains from stratification are greatest.”⁹ Given the high degree of correlation between facility size and the outcome variables for revenue and expenses, it is likely that stratification using this variable will reduce the variance of estimates. The initial sample will include 2,000 facilities sampled using Neyman allocation. Using past surveys, The Lewin Group has developed lognormal models for estimating costs and expenses within various modes of care. The independent variables include region, facility ownership type (private for-profit, private not-for-profit, and public), and client counts. We will use these models to estimate the variances within modes of care and facility types for the purposes of determining the number of facilities to be selected within each stratum. Separate models were developed for SA and MH facilities. The MH models were based upon the 2004 IMHO. The models for SA facilities were built using data from the 1998 Uniform Facility Data Set (UFDS) Survey, and applied to data from the 2006 N-SSATS. Separate models were developed for different modes of care, which result in different estimates of total annual revenue for each mode.

B2. Information Collection Procedures

a. Statistical Methodology for Stratification and Sample Selection

For stratification, we will employ both explicit and, by using sequential random sampling, implicit strata. Sequential random sampling controls the distribution of the sample by spreading it throughout the explicit stratum, based on the implicit stratification variables. The method of sequential random sampling that we plan to use is referred to as probability minimum replacement (PMR) as defined in Chromy (1979).¹⁰ The units on the frame will be sorted in a serpentine manner that maximizes proximity of similar units within explicit strata. “Similarity” is

⁸ Salvucci et al (2008), p. 15.

⁹ Cochran (1977), p. 127-128.

defined in terms of the variables selected for implicit stratification. Typically, as with other sampling schemes, certainty units with large variance are removed for separate handling to reduce the overall sampling error.

To reduce the variance even further, we will consider selecting the facilities with probability proportional to size (PPS) when the size indicator is known, using client counts as a size indicator. When size is not known, an imputed value will be used as a size indicator, or size will be set to a constant value (resulting in selection of facilities with equal probability).

b. Estimation Procedure

Plans for the statistical analyses of the data, is presented in Section A, at Question A16. SUDAAN will be used to provide standard error estimates to accommodate the sampling design.

Table 2 presents 95 percent confidence interval half-widths (half of the 95 percent confidence interval width) for the total revenue received as a percentage of the estimate of total revenue received, for all facilities and for subgroups of facilities defined by facility type. In this table, the expected sample size for each subgroup of facilities is determined by the proportion of the type of facilities in the universe.¹¹ Consider the subpopulation defined as “Psychiatric and Substance Abuse Specialty Hospitals.” The half-width for a 95 percent confidence interval for this estimate is 7.2 percent of the mean. The results reported are a percentage of the estimate, using a coefficient of variation of 20 percent. (We used this value because the coefficients of variation for total revenue received by health-related subpopulations in the Service Annual Survey¹² included a variety of published values, some of which exceeded 20 percent.)

No specialized sampling procedures will be used to accommodate unusual problems, and periodic data collection is not required since this is a one-time data collection.

b. Data Collection Overview

SAMHSA expects that the SSR&E will achieve at least a 75 percent response rate, based on the plan outlined here. The sampled facilities are in the business of treating people with substance abuse problems and mental illness; responding to surveys is not high on their list of priorities. Further, some of the information being sought is complex and may not be easily accessible at the facility level. Studies such as N-SSATS and NSMHTF once asked questions about revenues and expenses, but the additional burden these questions placed on facilities led to their being removed from the surveys. Thus, SAMHSA is taking a proactive approach to maximize the response rate for the SSR&E. First, SAMHSA is developing a marketing campaign prior to the survey. The campaign includes soliciting support from well-respected persons (such as the SSR&E Technical Expert Panel) as well as key associations such as the National Association of State Alcohol and Substance Abuse Directors (NASADAD), and the National

¹⁰ Chromy, J.R. (1979), “Sequential Sample Selection Methods,” 1979 Proceedings of the American Statistical Association, 401-406. Chromy’s method yields a sample of exactly the size wanted, one from each zone and a sample for which a closed expression for estimation variances can be calculated. After stratification, the sample will be selected using the probability-minimum-replacement, sequential sampling procedure developed by Chromy. Chromy’s procedure offers all the advantages of the systematic sampling approach but eliminates the risk of systematic, list-order bias by making independent selections within each of the zones associated with systematic sampling, while controlling the selection opportunities for units crossing zone boundaries. As a result, exact, unbiased expressions for the variance are available for Chromy’s procedure.

¹¹ Sample allocation based on revenue will improve the precision for the study sample.

¹² The Service Annual Survey is an annual survey conducted by the U.S. Bureau of the Census. It provides estimates of revenue and other measures for most traditional service industries. A new sample is introduced about every five years.

Association of State Mental Health Program Directors (NASMHPD). SAMHSA contractors will prepare marketing materials such as an explanation of the importance of the survey, frequently asked questions, and benefits of participation to the government and policymakers. Members of SAMHSA and the contractor team will take the opportunity to discuss SSR&E at relevant conferences and meetings. SAMHSA will ask individual TEP members to encourage facilities or organizations that are reluctant to participate. When SAMHSA and TEP staffs attend MH and SA conferences, they will be asked to distribute the campaign materials and encourage facility staff attending the meetings to participate. SAMHSA will ask N-SSATS and NMHSS project officers to insert a notification of the upcoming SSR&E into their mailing materials and on their websites.

Next, SAMHSA's contractors will develop training materials for staff. This will enable them to make advance calls to gain cooperation and help locate and identify the best facility respondent or respondents. A second training will be conducted for those who will later follow up with the facilities to see if they need assistance in responding or to retrieve any missing or unclear information. The follow up prompting calls and data retrieval are discussed more fully in Question B3 below. SAMHSA's strategy is to offer respondents either a web or paper option for completing the instrument as well as any assistance they might need from the facility liaison.

Collecting financial data from facilities is challenging: the data are sometimes proprietary and collecting them often requires "going up the ladder" to obtain data or permission from a corporate authority to release it. In addition, facilities and organizations collect and store their financial information in a variety of ways and many use non-federal fiscal years. This variability is one of the greatest obstacles facilities experience in providing financial information. Deciding how to fit their financial information into the categories used in the survey is a strong disincentive to respond.

As soon as OMB clearance is obtained, SAMHSA's contractors will distribute an Invitation to Participate packet to the 2,000 sampled MH/SA facilities using unduplicated contact information from the N-SSATS and NMHSS. The materials in the packet will be customized to the MH or SA facilities. The packet will include a letter from SAMHSA explaining the survey and its importance to SAMHSA and the research community at large, introducing the contractor team. It also will include a letter from relevant local organizations (NASADAD, NASMHPD) supporting the survey and encouraging participation. Finally, the packet will contain a SSR&E Brochure that will present key information about the survey (including its confidentiality) and will address questions that we expect to be asked based on the pilot test outcomes. Attachment B includes the Invitation to Participate.

The contractor's facility liaison staff will begin advance calling after the OMB has approved data collection and the sample is selected. At that time, the liaison will telephone the facility contact to confirm receipt of the packet (which will be re-mailed if necessary) and to work with the contact to identify the best respondent at either the facility level or the organizational level. If it is necessary to obtain approval for participation from another authority higher in the organization, the facility liaison will mail that individual a set of materials similar to the notification packet requesting approval for participation. The facility liaison will call to verify that the packet was received and will negotiate approval for the facility to participate.

SAMHSA contractors will begin conducting the survey follow-up calls approximately one month after OMB clearance. The calls will encourage facility participation, assist facility staff in providing information in a way that is easy for them and acceptable to the research design, and retrieve missing or erroneous information. If necessary, liaisons can help the facility respondent with the web instrument or answering by paper, or will even work with the individual to respond by telephone.

In addition, data editing and validation will begin as soon as data collection starts. SAMHSA will specify what data need to be retrieved: all missing data or only missing critical items. On a flow basis, the facility liaisons will inform the team of any missing information or errors found. The liaisons will immediately act upon this information and seek to retrieve the data SAMHSA requires through subsequent phone calls. In addition, contractor staff will review all data, web or paper, for information that appears to be incomplete or contradictory. They will contact the facility to clarify the incomplete or contradictory information.

Once a facility has responded, the contractor will send a letter thanking the facility for their help.

Table 5 shows the approximate schedule of survey events. While the start of data collection will depend on receiving OMB clearance, the intervals between tasks should remain the same.

Table 2. Planned Survey Schedule

Activity	Start Date	End Date
Marketing campaign	01/06/09	02/05/10
Implement instrument in WebSurv	1/08/10	2/15/10
Train for advance calls	2/15/10	2/16/10
Train for follow-up calls	3/15/10	03/16/10
Begin data collection	2/17/09	8/17/09
Coordinate mailing and calls	2/17/09	04/23/10
Conduct follow-up calls	4/23/10	08/17/10
Maintain tracking system	2/17/10	8/17/10
Clean and prepare analytic files	06/01/10	08/26/10

B3. Methods to Maximize Response Rates

a. Maximizing Response Rates

SAMHSA expects the SSR&E to achieve at least a 75 percent response rate, based on the proactive outreach and recruiting plan outlined above (B2c), contractor experience conducting other SAMHSA surveys and establishment surveys, and carefully crafted follow up procedures. Contractor staff will begin conducting the survey follow up calls approximately six weeks after the recruiting calls begin. While the primary purpose of the follow up calls will be to encourage participation, the callers will also offer to assist facilities in any way needed to provide SSR&E information. To make responding easy, facilities may choose to respond using a paper instrument, a web instrument, or by having contractor staff assist them.

During data collection, the tracking system will identify nonresponders and the facility liaisons will follow up to encourage participation and offer assistance. The survey team will review frequencies early and often to swiftly identify and retrieve missing, nonresponse, or out-of-range data. If the data are submitted by mail or fax, they will be keyed into the web instrument and the frequencies checked in the same way as web-collected data. MPR will track response rates by stratum and focus on obtaining equivalent response rates throughout the strata.

b. Handling Nonresponse

All statistical analyses with data from the SSR&E will require procedures that accommodate different types of nonresponse in the data. Design weights are calculated by taking the inverse of the probability of selection. The procedures described below will account for two types of nonresponse: unit nonresponse, in which the entire observation unit is missing, and item

nonresponse, in which some measurements are present for the observation unit, but at least one item is missing.

Unit Nonresponse. Unit nonresponse is usually alleviated by the calculation of nonresponse weights. A commonly used method to compute nonresponse weights is to form classes of sample members with similar characteristics, and use the inverse of the class response rate as the adjustment factor in that class. “Weighting classes” are formed to ensure there are sufficient counts in each class to make the adjustment more stable (that is, to have a smaller variance). The natural extension to the weighting class procedure is to use logistic regression with the weighting class definitions used as covariates, provided each level of the model covariates has a sufficient number of sample members to ensure a stable adjustment. The logistic regression approach also has the ability to include both continuous and categorical variables, and standard statistical tests are available to evaluate the selection of variables for the model. The nonresponse weight is then determined by grouping the predicted probabilities of response into weighting classes and taking the inverse of the class weighted response rate. The final analysis weight is the product of the design weight and the nonresponse weight.

Since stratification will be used when selecting the sample of facilities, the assumption of a simple random sample (SRS) that is required in most of the traditional methods for calculating variances will not be satisfied. For non-SRS samples, we can take repeated subsamples from the main sample and use these subsamples to estimate the variance. This is done by recalculating the weights for each subsample, calculating the estimates in each subsample using the recalculated “replicate” weights, then calculating the variance of the estimates. There are a variety of methods available for drawing these random subsamples, including Balanced Repeated Replication (BRR), Random Groups, and Jackknife. MPR has experience with all of these methods and will implement the method that is most appropriate for the final design.

Item Nonresponse. We expect to deal with item nonresponse by employing imputation. In this survey, three general approaches are proposed: (1) logical imputation, in which unreported data elements can be determined logically from reported data elements (for example, client counts are reported by category but total clients are unreported); (2) cold deck imputation, in which existing surveys can be used to impute data elements using data reported in the N-SSATS or NSMHTF (for example, client counts by type of service, facility type, and ownership status); and (3) model-based imputation.

Model-based approaches assume that variables with missing values are functionally related to other variables that have existing, reported data. An example of a model-based approach is the Estimation-Maximization (EM) method. (Details on the EM algorithm are found in Little and Rubin 2002.) The intuition underlying the EM algorithm, which uses a maximum likelihood approach, follows. If we know the parameters of a maximum likelihood model describing the relationship between the dependent variable and the explanatory variables—the beta coefficients—we could estimate the value of any given data element by choosing that value for which the likelihood value is a maximum. Similarly, if we have the data for all the explanatory variables and the dependent variables, we can estimate the coefficients of the model using maximum likelihood.

The EM algorithm improves on traditional methods of imputing missing values, such as hot deck or regression imputation. In hot deck imputation, random selection of the data element is made from a set of observations that have the same or similar characteristics as the observation with the missing data element. A drawback of this method is that relationships between the imputed value and reported values for other survey data elements may be unreasonable, depending on the parameters of the hot deck function. Regression methods predict the missing value using a regression of the variable in question on other variables. It has an advantage over

the hot deck methods in that it more systematically conditions the estimate of the missing data element on the values of the elements that are available, accounting for the relationship between the missing data element for the observation and the other elements for that observation. A drawback to the regression approach is that it does not address variance across facilities with the same characteristics in the model.

In an initial step, the maximum likelihood values of the parameters are estimated using available data. Then the model is used to obtain values for the data items that have missing data, using the initial estimate of the coefficients. Given the values for the missing data elements, the coefficients are estimated again. A new set of predicted missing values is then estimated. The algorithm is repeated until the changes in the estimated coefficients (and the value of the missing data elements) are satisfactorily stable. Attachment A contains a copy of the survey instrument.

The survey will include the following data elements:

- All sources of revenue, with specification of dollar amounts
- Total aggregate and sub-aggregate expenses, including operational and utility costs, other contracted services, depreciation and capital expenses.
- Client counts, and mix of clients between those with mental illness, substance use disorders, or both.

For each of the variables requiring imputation, appropriate dependent variables will be selected for the model. The covariates for the imputation models will focus on such commonly reported variables as client counts and ownership; however, for subsets of facilities with more complete reporting, the functional form would be expanded to include less frequently reported variables that are highly predictive. For example, for facilities that report expenses but not revenue, the reported expenses would be used to predict revenues. However, for facilities that report neither revenues nor expenses, client counts, ownership and other more commonly reported variables will be used.

Nonresponse Bias Analysis. As in any survey, some sampled facilities in the SSR&E will not be able or willing to respond to the survey. These unit nonrespondents have the potential of causing nonresponse bias in estimate, but this does not necessarily imply that survey estimates will exhibit bias.

The purpose of the nonresponse bias analysis is to provide some indication of whether a possible nonresponse bias does exist, an indication of the data items and populations for which survey estimates may have a greater potential for bias, and the possible extent of nonresponse bias in survey estimates. However, because survey data is not generally available for nonrespondents, we can never be certain if bias does or does not exist in the survey estimates.

To evaluate the bias resulting from nonresponse in the SAMHSA Survey of Revenues and Expenses (SSR&E), SAMHSA will compare attributes of respondents and nonrespondents by looking at data available for both from the sample frame and external survey data.

Assessing Nonresponse Bias Using Data from the Sample Frame and from Other Surveys. This type of nonresponse bias analysis will use the data in the sampling frame that are available for respondents and nonrespondents. These include characteristics about the facilities that are available on the sample frame (geographic location, number of facilities per organization, primary focus of care, and mode of care). Other characteristics will be available for many nonrespondents, but are not available for large subsets of nonrespondents among MH facilities. These variables are facility type, number of beds, client counts, ownership type, affiliation, and specific attributes of treatment in the facility. The nonresponse bias analysis also will use estimates of revenue from models developed from using organization-level MH data in

the 2004 Inventory of Mental Health Organizations (IMHO) and facility-level data in the 1998 Uniform Facility Data Set (UFDS). These models will be applied to data from the 2008 NSMHTF and 2007 N-SSATS.

For the nonresponse bias analysis using frame and external survey data, SAMHSA intends to:

- Compute response rates for key subgroups of MH and SA facilities.
- Compare the weighted distributions of respondents and nonrespondents for estimated revenues using models based on data from previous surveys.
- Identify the characteristics that best help predict nonresponse through a Chi-square automatic interaction detection (CHAID) analysis and logistic regression modeling, and use this information to generate nonresponse weight adjustments.
- Compare the distributions of respondents using the fully response-adjusted analysis weights for sampling frame characteristics to the distributions for the main sample comparably weighted using the unadjusted sampling weights.

These analyses will be conducted within and across strata to assess whether the potential for nonresponse bias differs across strata. Below, we discuss each of these steps in greater detail.

Compute Response Rates for Subgroups. The response rates will be computed using the AAPOR definition of the response rate—that is, the weighted number of completed interviews with eligible facilities divided by the estimated number of eligible facilities.¹³

Response rates will be computed overall and for subgroups to examine if these differ systematically. We will compute four measures of differences in subgroup response rates relative to overall response rates:

1. Simple difference: Rate for a specific category—overall rate
2. Absolute difference: the absolute value of the simple difference (Rate for a specific category—overall rate)
3. Relative simple difference: the simple difference divided by the overall rate ([Rate for a specific category—overall rate] / overall rate)
4. Relative absolute difference: the absolute difference divided by the overall rate ([Rate for a specific category—overall rate] / overall rate)

We will review these measures and describe the patterns in nonresponse. This analysis will only assess the response patterns as simple “main effects.” The third step will assess potential interactions in response patterns among subgroups.

Compare the Characteristics of Respondents and Nonrespondents. Next, SAMHSA will examine the distributions of respondents and nonrespondents for estimated revenue using data from previous surveys. The Lewin Group has developed lognormal models for estimating revenues within various modes of care. The independent variables included region, facility ownership type, and client counts. SAMHSA will apply these models to estimate revenues using data from the 2008 NSMHTF and 2007 N-SSATS, using separate models for different modes of

¹³ The American Association for Public Opinion Research. 2008. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Fifth edition. Lenexa, Kansas:

care within SA and MH facilities. Estimates will be generated using the initial (sampling) weights for nonrespondents and respondents. This type of analysis can be useful in identifying patterns of potential nonresponse bias, though several sources of variance make it difficult to attach tests of significances to any differences that occur.

Identify the Best Explanatory Factors of Nonresponse and Generate Nonresponse Weight Adjustments. Logistic regression modeling is commonly used to develop adjustment factors for nonresponse, also known as response propensity modeling. Response propensity modeling using logistic regression can be viewed as an extension of the classical weighting-class nonresponse adjustment procedure that makes it possible to include more factors (that is, binary, categorical, and continuous factors) in nonresponse adjustments. To simplify the process, Chi-square automatic interaction detection (CHAID) is commonly used to assist in identifying potentially significant interactions among the subgroups or factors available for all individuals. SAMHSA plans to use CHAID, with the initial sampling weights, to help identify the interactions in a multiple pass process.

The CHAID algorithm partitions the sample in a hierarchical fashion, with each successive splitting of the sample identified by CHAID. CHAID uses the Chi-square statistic with the proportion responding defined as the dependent variable to determine the partitioning of the sample with the largest value for the statistic among all possible partitions by the factors available. After the initial partitioning, the Chi-square statistic is again used to identify additional partitions subject to pre-determined restrictions (for example, a minimum partition size).

Because such “hierarchical” splitting can miss potentially important interactions, after the first CHAID analysis, we remove the initial “branching” variable and rerun the CHAID algorithm with this variable excluded. If the CHAID analysis reveals the same basic branching pattern for response rates, we proceed to the logistic modeling step (described below). If not, we remove the initial branching variable for the second CHAID analysis and rerun the CHAID algorithm a third time. Our experience is that three CHAID steps are sufficient to identify the most important interaction terms.

Next, we develop variables that reflect the interaction terms identified through the CHAID analyses, and use these variables in forward and backward stepwise logistic regressions to eliminate redundant interaction variables and to identify the most significant interactions. The stepwise logistic regressions are conducted using SAS software with normalized weights. However, the SAS software for stepwise logistic regression does not account for the sampling design. Hence, we use SUDAAN to develop the final model, so variance estimates for the coefficients reflect the sampling design. Goodness-of-fit for the final model is assessed using the percentage of concordance and discordance, the R-square for the model, and the Hosmer-Lemeshow Goodness-of-fit test statistic.

The final response propensity model described above will be used to identify factors associated with nonresponse and to compute the appropriate nonresponse adjustment factors for the sampling weights. The inverse of the predicted propensity to respond will be used as an adjustment factor to the initial sampling weights. These response-adjusted weights will then be post-stratified to baseline marginal totals for the main study population and will be the final analysis weights.

Compare the Fully Adjusted Weighted Distributions of Respondents Along Baseline Characteristics to the Distributions for the Main Sample. In this last step, SAMHSA will generate estimates of the distribution of respondents along sampling frame characteristics using the fully adjusted analysis weights and compare these distributions to the known totals for the main study population and for key subgroups. Analogous to the assessment of response rates, SAMHSA will compute four measures of differences relative to the main sample:

- Simple difference: weighted estimate for respondents–frame total for a specific category.
- Absolute difference: the absolute value of the simple difference (weighted estimate for respondents–frame total for a specific category).
- Relative simple difference: the simple difference divided by the frame total ([weighted estimate for respondents–frame total for a specific category] / Frame total).
- Relative absolute difference: the absolute difference divided by the overall rate ([weighted estimate for respondents–frame total for a specific category] / Frame total).

This analysis can highlight measures where the potential for nonresponse bias is greatest and where greater caution should be exercised in the interpretation of the observed findings.

In summary, the availability of data from the sampling frames (including N-SSATS and NSMHTF survey data) as well as estimated revenues from previous surveys, will allow an in-depth assessment of any potential for nonresponse bias and the estimates that may be affected. SAMHSA will use the results of this nonresponse bias analysis in the preparation of the reports to highlight substantive topics that are unlikely to be affected by nonresponse bias and to provide appropriate cautionary statements for findings that may be vulnerable to nonresponse bias.

SAMHSA is exploring a second possibility for detecting nonresponse bias but has not yet funded it. If funding becomes available, SAMHSA will assess the effect of nonresponse bias on estimates using a subsample of targeted sampled cases.

Assessing Effect of Nonresponse Bias on Estimates Using Sampled Cases Only. SAMHSA will target a small subsample of facilities among the 2,000 sampled facilities. In this subsample, nonrespondents will be targeted with extraordinary efforts to ensure no nonresponse. It will act as a control subsample, which, if we are successful in obtaining close to 100% response, will have no nonresponse bias.¹⁴ The rest of the sample will entail another independent subsample, which we will call the “main study subsample.” It will have an assumed 75% response rate, where no extra effort will be induced for nonresponders.¹⁵ Estimates of total net revenue from this control subsample will be compared with estimates from the main study subsample. Facility size measures and services provided also will be compared, as will revenue estimates within subpopulations defined by these variables and mode of care, which is available on the sample frame.

Data collection would occur for the control subsample facilities in the same manner as the facilities in the main study subsample, and would start at the beginning of the data collection timeline for the survey. The size of the control subsample will depend upon the cost that SAMHSA is willing to incur, and the power that each option provides. The sample of 2,000 facilities will be released in waves, with the facilities in the control subsample concentrated in the first wave. For example, if the control subsample comprises 100 facilities, we would expect responses from 75 of these 100 facilities (a response rate of 75 percent) and expect 25

¹⁴ It is possible that the responses from the initial nonrespondents in this subsample (i.e., prior to intensive followup procedures) may systematically differ from the responses of the extra-effort respondents. (For example, extra-effort respondents may put less effort into their responses, making them less reliable.) For the purposes of this study, we assume this measurement error bias is zero.

¹⁵ We separate the samples into two subsamples for simplicity, to ensure independence of the two samples.

nonrespondents among these 100 facilities, for which the intensive follow-up will occur.¹⁶ This intensive follow-up will include phone interviews of 20 of the control subsample nonrespondents, with the phone interviews performed by a survey specialist instead of a standard interviewer. For the remaining 5 control subsample nonrespondents, responses will be obtained from in-person visits to the site. For the remaining 1,900 facilities in the main study subsample, a 75% response rate will result in 1,425 respondents.

To facilitate 100 percent response from facilities in the control subsample, the questions from the questionnaire that will be asked of the nonrespondents in the control subsample will be limited to critical items in the questionnaire, in four general categories: screening information (all questions in Section A), services provided (questions B1-B5), total net revenue (question C1), and size of facility (questions E0, E1a, E1b, E1c, E6a, E6b, and E6c).

The proposed sample design for the entire sample has 52 strata; therefore, it would not be possible to select a control subsample of 100 cases using all strata. Because different response patterns may exist across the sampling strata, we will incorporate stratification into the subsample design, collapsing the 52 strata to ensure a representative sample. Because mode of care and facility size are the most important stratification variables, we will proportionally allocate these 100 facilities within strata defined by mode of care and size, and select the subsample of facilities independently within each stratum.

Two sets of weights will be created. For each subsample, analysis weights will be calculated in the same way: they will be created as a product of the probability of selection (based upon the original sample design), the probability of selection into the subsample, and a nonresponse adjustment. No nonresponse adjustment will be required for the control subsample if a 100% response rate is actually achieved. Weighted estimates of total net revenue, facility size measures, and services provided will be compared between the two samples. In addition, estimates within subpopulations defined by facility size and services provided, as well as mode of care, will also be compared. Table 6 provides estimates of the minimum detectable difference as a proportion of the mean revenue. Coefficients of variation (CVs) ranging from 0.3 to 1.0 are presented, though work with data in the National Survey of Substance Abuse Treatment Services (N-SSATS) indicates that a CV of 0.3 to 0.5, obtainable through stratification by facility size, is more realistic than a CV of 1.0.

Table 6. Minimum Detectable Differences Comparing Control Subsample Estimates with Main Study Subsample Estimates, In Terms Of Proportion of the Mean Revenue

	CV=0.3	CV=0.5	CV=1.0
Main study subsample	11.5%	19.2%	38.3%
50% subpopulation	16.3%	27.1%	54.2%

Note: For the MDD calculations, we assumed: (1) a 75% response rate in the main study subsample; (2) 80% power and a 5% level of significance; (3) a design effect of 1.75, due to unequal weighting; (4) a control subsample of 100 respondents (100% response rate); and (5) a main study subsample with 1425 respondents (75% response rate).

As Table 6 indicates, differences will need to be fairly large to be detectable, particularly if comparisons are made within subpopulations.

¹⁶ A lower response rate is possible. In that instance the number of nonrespondents for whom extra effort is required will necessarily be larger and the necessity of this study even more apparent.

In many nonresponse bias analyses, data from the sampling frame and from external sources are used to evaluate the differences between respondents and nonrespondents. If necessary, estimates are adjusted using information from these analyses. We will be doing this type of analysis for the SSR&E (although we don't anticipate using the results of the analysis to adjust estimates). Additionally, if survey data can be obtained from some nonrespondents through extraordinary follow-up efforts, attributes from this survey of initial respondents and initial nonrespondents can be compared to assess nonresponse bias. Methods for subsampling nonrespondents for further follow-up to obtain survey data after the completion of initial data collection are described in Cochran (1977) and Sarndel, Swenson, and Wretman (1992). Although that is not the primary purpose of this exercise, respondents and nonrespondents can be compared using these data, provided a larger control subsample is employed. Table 7 presents a comparison of minimum detectable differences using the control subsample as it is currently defined (with 25 expected nonrespondents prior to intensive follow-up) and a control subsample with 400 sample members (with 100 expected nonrespondents prior to intensive follow-up).

Table 7. Minimum Detectable Differences Comparing Estimates from Control Subsample Nonrespondents Prior To Intensive Follow-Up and Main Study Subsample Respondents, Assuming A Control Subsample Size of 100 And 400, As A Percent Of Mean Revenue

	CV=0.3	CV=0.5	CV=1.0
Control subsample size = 100			
All followed up cases (n = 25)	22.4%	37.4%	74.7%
50% subgroup (n = 12)	32.4%	53.9%	107.8%
Control subsample size = 400			
All followed up cases (n = 100)	11.5%	19.1%	38.3%
50% subgroup (n = 50)	16.2%	27.1%	54.1%

Note: For the MDD calculations, we assumed: (1) a 75% response rate in the main study subsample; (2) 80% power and a 5% level of significance; and (3) a design effect of 1.75, due to unequal weighting.

B4. Tests of Procedures

SAMHSA conducted a two-phase pilot test of the survey instrument with a purposive sample of nine MH and/or SA facilities. SAMHSA made numerous changes and deletions to the SSR&E instrument based on the results (as well as comments from the public at the end of the 60-day comment period). The Phase 1 Pilot Test instrument (conducted with six facilities) contained a complex section on facility staffing and staff turnover. The questions in this section proved extremely difficult for facilities to answer in a timely manner especially as each facility retained staffing information in different not easily accessible formats (see the Report for detail). The staffing and staff turnover sections were deleted from the Phase 2 instrument (administered with three additional facilities). The Phase 2 instrument was more streamlined and certain definitions and questions were further refined. SAMHSA needed to make very few changes to the Phase 2 instrument – with the exception of complying with some public comments that helped clarify question and definition language. Attachment C contains the Pilot Test Report, including detailed discussions of issues raised during both phases of the Pilot Test as well as copies of the two versions of the Pilot Test instrument. The Pilot Test Report identifies specific recommendations that SAMHSA has accepted. Attachment A contains the final version of the SSR&E instrument that incorporates all Pilot Test and public comment suggestions. Further, as a result of the pilot test, SAMHSA decided to structure the topics to be covered in the advance call. The introductory topics are found in Attachment G.

B5. Statistical Consultants

Mathematica Policy Research, Inc., 600 Maryland Ave., S.W., Washington, D.C. 20024, (202) 484-4215

James Kretz, Senior Survey Statistician and Project Officer, SAMHSA, 1 Choke Cherry Road, Rockville, MD 20857, (240)-276-1755.

**ATTACHMENT A
SAMHSA SURVEY OF REVENUES AND EXPENSES (SSR&E) QUESTIONNAIRE**

**ATTACHMENT B
INVITATION TO PARTICIPATE PACKET SAMHSA LETTER OF INVITATION
FACT SHEET ABOUT THE SURVEY THANK YOU FOR PARTICIPATING LETTER**

**ATTACHMENT C
PILOT TEST REPORT PHASE 1 INSTRUMENT (MAY 13, 2009) PHASE 2
INSTRUMENT (AUGUST 3, 2009)**

**ATTACHMENT D
SURVEY OF REVENUES AND EXPENDITURES QUESTION SOURCES**

**ATTACHMENT E
RESPONSES TO PUBLIC COMMENTS**

**ATTACHMENT F
TECHNICAL ADVISORY PANEL MEMBERS**

**ATTACHMENT G
INTRODUCTORY TOPICS**