

2003

The Methodology of the Multi-site Study of the Termination of Supplemental Security Income Benefits for Drug Addicts and Alcoholics

James A. Swartz

Peggy Tonkin

Jim Baumohl

Bryn Mawr College, jbaumohl@brynmawr.edu

[Let us know how access to this document benefits you.](#)

Follow this and additional works at: http://repository.brynmawr.edu/gsswsr_pubs

 Part of the [Social Work Commons](#)

Custom Citation

Swartz, James, Peggy Tonkin, and Jim Baumohl. "The Methodology of the Multi-site Study of the Termination of Supplemental Security Income Benefits for Drug Addicts and Alcoholics." *Contemporary Drug Problems* 30, no. 1/2 (2003): 77-121.

This paper is posted at Scholarship, Research, and Creative Work at Bryn Mawr College. http://repository.brynmawr.edu/gsswsr_pubs/49

For more information, please contact repository@brynmawr.edu.

The methodology of the multi-site study of the termination of Supplemental Security Income benefits for drug addicts and alcoholics

BY JAMES SWARTZ, PEGGY TONKIN,
AND JIM BAUMOHL

This paper describes the quantitative and qualitative methodologies used in a nine-site, two-year study of the effects of terminating Supplemental Security Income (SSI) for drug addiction and alcoholism (DA&A). The quantitative component of the study involved a longitudinal survey that collected data on 1,744 former DA&A recipients, representing about one-fourth of the national population, and achieved an aggregate follow-up rate of 82%. Despite limitations in questionnaire design and implementation, the survey provided reasonably valid data in the following areas: demographics, employment/income, medical/psychiatric status, drug and alcohol use, legal involvement, family/social functioning, food and hunger, housing, and victimization. The qualitative component examined the lives of a subsample to help clarify important issues that could not be addressed within the more structured protocol and format of the longitudinal survey. The paper also presents details on the survey instrument design, the results of validation studies of selected survey items, and data collection protocols across study sites.

In March 1996 federal legislation (P.L. 104-121) rescinded Supplemental Security Income (SSI) and Social Security Disability Insurance (DI) benefits to recipients with disabling impairments “materially related” to drug addiction or alcoholism. Benefits based on these impairments ceased as of January 1, 1997. This paper details the methodology employed in a multi-site longitudinal panel study (the SSI Study) that aimed to evaluate the impact of this policy change on former recipients, known as drug addiction and alcoholism (DA&A) beneficiaries.

The SSI Study developed opportunistically, growing out of two ongoing programs at the Center for Substance Abuse Treatment (CSAT) and two studies proposed to the Robert Wood Johnson Foundation (RWJ). The first CSAT project was the Target Cities Demonstration Program. This sought to centralize intake and assessment, allow for treatment matching, and effect a number of other systems changes in major metropolitan areas. Its principal goal was to improve both access to treatment and services for substance abusers. A number of the Target Cities sites (San Francisco, Los Angeles, Portland, Chicago and Detroit) had enrolled significant numbers of DA&A recipients. The second CSAT program was a demonstration project funded by the Social Security Administration (SSA) through CSAT to test the effectiveness of intensive case management for DA&A recipients. There were two active projects at the time of the DA&A policy change: one in King County (Seattle), Washington, and the other in Detroit.

During the summer of 1996, in consultation with the SSA, CSAT was in the process of determining how to proceed with the King County and Detroit projects, given that the DA&A program had been terminated. At the same time, CSAT became aware of two proposed studies of the impact of the DA&A policy change that RWJ was likely to fund. These were in Chicago and in Santa Clara (San Jose) and San Joaquin (Stockton) counties in Northern California. Together,

it seemed that the CSAT programs and the proposed RWJ sites could form the basis of a multi-site study. CSAT's interest in studying DA&A recipients derived from the assumption that members of this population were heavily involved with alcohol and other drugs and that removal of the federal mandate that they participate in treatment (along with their loss of monitored cash benefits) had potentially adverse consequences (e.g., increased drinking and drug use, lower participation rates in treatment, health problems related to increased drug use).

In the late summer and early fall, the CSAT and RWJ sites forged a collaboration, with CSAT providing funds to increase the sample sizes at the RWJ sites and to extend the study period from one year to two years. Table 1 lists the sites that participated in the SSI Study and the geographic coverage for each. Since the University of Akron already acted as the data-coordinating center for the Target Cities program, participants in the SSI Study agreed that since Akron had the expertise and resources to coordinate the proposed study, it would be practical and cost effective to continue this arrangement.

TABLE 1

Geographic coverage by SSI study site

Site Name	Geographic Coverage
Chicago	Chicago metropolitan area only
Detroit	Detroit metropolitan area only
Seattle	King County, including metropolitan Seattle
Portland	All of Multnomah, Washington, and Clackamas counties, including metropolitan Portland
Oakland	All of Alameda County, including metropolitan Oakland
San Jose	All of Santa Clara County, including metropolitan San Jose
Stockton	All of San Joaquin County, including metropolitan Stockton
San Francisco ¹	All of the co-terminus city and county
Los Angeles	All of Los Angeles County, including metropolitan Los Angeles

The DA&A population encompassed by the SSI Study *does not* represent the SSI DA&A population nationally (see Wittenburg et al., this issue). For example, there were no sites from the South or the East Coast. Collectively, however, the sites *do* reflect a demographic cross-section of the national DA&A population, as well as a range of ecological factors; there was wide variation among the sites in terms of the availability of social services, including substance abuse treatment and funding for things like housing subsidies, medical coverage, and state and/or local welfare. There were also significant differences in general population characteristics and employment opportunities. Although funding expedience and other exigencies played substantial roles in site selection, CSAT selected many of the sites because of their relatively large number of DA&A recipients. As of March 1, 1996, the sampling frames of the nine study sites represented about 26% of the national DA&A population (Wittenburg et al., this issue).

Because there was considerable variation across sites in available social benefits, the study intended to examine whether different levels of local support mediated the effects (if any) of the changes in federal legislation. Thus, despite the limitation of not having a nationally representative sample, CSAT officials thought that the diversity of the subjects and contexts would allow for more confidence in the findings should there be consistency in the cross-site results and that the results would illuminate specific areas for targeting additional drug treatment and support services.

Study design

The collaborators had two months to set in motion perhaps the most ambitious study of welfare reform conducted to date. In October 1996 CSAT convened the first meeting of the SSI Study Group, comprised mainly of the primary inves-

tigative staff from each site as well as a few invited experts in the areas of welfare and disability policy study. At this meeting and at many subsequent gatherings, the Study Group developed a set of questions that were incorporated into a cross-site data collection instrument as well as a standard interview protocol to enable valid cross-site analyses. The intent of these analyses would be to determine which effects (if any) of the federal policy change could be generalized across sites, and which effects were site-specific and might be attributable to confluences of local factors and subject differences. Most of the papers contained in this issue utilize the cross-site data set that resulted from these efforts to study the general and specific effects of the legislation.

CSAT officials sought to provide project oversight and arbitration where investigators were in dispute over design issues, but the study was conducted as a multi-site collaboration rather than as a centrally administered project. CSAT had authority only with respect to the sites that it funded. The RWJ sites were free to collaborate within the goals of their studies. The collaborative structure provided considerable flexibility in allowing sites to supplement the survey questionnaire in order to address research issues of local concern, to have final say in sample size, and to develop supporting materials such as consent forms to meet the needs of local authorities. This flexibility in turn led to the development of more locally meaningful research projects and the study of issues not covered in the core survey instrument (e.g., psychiatric diagnoses were assessed in Chicago; a standardized questionnaire on hunger was administered at the Northern California sites). However, the benefits of this flexibility were offset to an extent by variations in sampling frames, recruitment techniques, retention rates, and questionnaire administration that potentially limit the consistency and generalizability of the composite data set. We will discuss these issues in detail later.

We should note that the SSI Study actually consisted of two complementary investigations. Nine sites conducted a longitudinal survey that used large samples, multiple interview waves, and a structured interview format. The longitudinal survey produced the composite data set referred to above. In conjunction with the longitudinal survey, four of the sites participated in a more limited qualitative study. The design and intent of the two studies were quite different.

The longitudinal survey was designed as a two-year panel study during which large groups of DA&A recipients would be interviewed five times—at baseline and at four follow-up interviews at six-month intervals. Because some DA&A beneficiaries were expected to requalify for benefits under a different impairment category (see Hunt and Baumohl, this issue), the longitudinal survey approximated a natural field experiment whereby those who requalified for SSI benefits could be treated as a comparison group. Subjects who retained SSI benefits could be compared on a variety of indicators with subjects who lost benefits. The Study Group hypothesized that subjects who lost benefits would generally fare more poorly than those who did not, but that differences might diminish as subjects adjusted to the loss of benefits.

The supplemental, qualitative study was conceived and developed after collection of the second wave of survey data. At that time, preliminary analyses revealed limitations in the survey data on key research issues. For example, the many subjects who reported little or no income also reported having stable housing and consistent sources of food. Also, the survey data did not allow for examination of *individual patterns* of change over time, such as how a person might move between jobs or treatment programs *within* one of the interview periods, nor did it allow for an in-depth understanding of how subjects understood SSI regulations and procedures, their reasons for deciding to reapply (or not) for disability benefits under a different impairment category, or their reasons for seeking disability benefits in the first place. The

qualitative study team developed a semi-structured interview instrument to address these issues. The design called for the collection of data in only four sites; used a semi-structured, modified life-history interview; used a targeted subset of the subjects interviewed for the longitudinal survey; and was conducted only once—between the third and fourth waves of the survey study. Although not intended to be a validation study per se, the qualitative study helped investigators interpret the survey data and pointed out areas where the quantitative data might be suspect (e.g., self-reported income levels).

In the sections that follow, we describe in more detail the methodology of both the longitudinal survey and the supplemental, qualitative inquiry. In describing the longitudinal survey, we consider methodological aspects that were relatively invariant across sites (i.e., eligibility criteria, core instrument items, and data collection plan) and how the sites varied in terms of their sampling frames, recruitment rates, and other dimensions that may have affected the generalizability of the findings to the aggregate SSI population covered by the sites and to the comparability of the data across sites. Immediately following the description of the longitudinal survey, we describe the results of a host of ancillary studies undertaken to determine the validity of different content areas assessed by the core survey instrument. These studies were done because of the Study Group's concerns about using an instrument that had not been validated in prior work and because respondents were not reporting the number and magnitude of problems expected. Therefore the Study Group wanted to determine if responses in critical content areas (e.g., rates of drug use, requalification rates) were valid. Following this section, we describe the methodology used in the qualitative study. The paper concludes with a consideration of the overall strengths and weaknesses of the SSI Study given the methodological choices made by the Study Group and the variations in study implementation across sites.

Survey methodology

Eligible subjects DA&A recipients were eligible for the study if they were between the ages of 21 and 59 and were in active pay status for SSI DA&A benefits in 1996. However, each site created its own sampling frame and extracted data on DA&As at different times and from different sources. For example, in Seattle subjects were eligible if they were in active pay status as of October 1996, while in Chicago the sampling frame included those receiving DA&A benefits as of June 1996. Differences among sampling frames are discussed in more detail below.

We excluded recipients with legal guardians and those who received DI—so-called “concurrent beneficiaries.” (See Hunt and Baumohl, this issue, for a detailed description of the distinction between SSI and DI as well as how the two benefits could be received concurrently.) The study focused on SSI-only cases because, by the terms of program eligibility, these individuals had no recent work histories and very limited material assets. The Study Group reasoned that if ending the DA&A impairment category had adverse effects, these probably would show up first among the SSI-only group of recipients and have the most pronounced impact on them. All but the California sites excluded DA&A recipients residing in institutional settings such as prisons, jails, or psychiatric hospitals at the time of the baseline interview because of the additional cost and time involved in locating these subjects and gaining clearance to conduct interviews in the institutions. Each site arranged translators for the very small number of otherwise eligible respondents who did not feel comfortable doing the interview in English. The Northern California sites developed a Spanish version of the questionnaire.

Despite the SSI-only eligibility requirement, the qualitative research team discovered that some sites inadvertently recruited and interviewed concurrent SSI/DI beneficiaries in

their baseline samples. The data from these subjects were subsequently discarded so that the composite data set excluded all concurrent beneficiaries, though each site was free to retain data on these cases in its local data set.

Sampling frames

Prior to the first meeting of investigators from the study sites, CSAT attempted to obtain from the SSA its master lists of DA&A recipients within the geographic catchment areas of the study. Unfortunately, confidentiality considerations prevented the timely availability of these lists. Consequently, the study sites had to rely on two available sources of information to identify eligible subjects who resided in the geographic areas listed in Table 1. In seven sites, researchers obtained lists of DA&A recipients meeting the study's eligibility criteria from the State Referral and Monitoring Agencies (RMA) (see Hunt and Baumohl, this issue, for a description and history of the RMA). Two sites, Chicago and Los Angeles, secured lists of DA&A beneficiaries from their regional SSA offices. The two types of lists had significant differences. Most important, at a majority of the sites the RMA lists were not as complete as the SSA lists (see Wittenburg et al., this issue). Thus any site using the RMA list to define its sampling frame did not include all DA&A recipients eligible for the study. Moreover, those most likely excluded from the RMA lists were those hardest to contact or the least compliant with the SSA's requirements for maintaining DA&A benefits (particularly the requirement to participate in treatment when deemed appropriate). Thus there is a good possibility that DA&A beneficiaries excluded from the sampling frames at sites using RMA lists differed in some systematic way from those not excluded, though the ultimate impact of these differences on study findings is hard to gauge.

Estimates of the coverage rates of the RMA lists range from about 48% in Portland to as high as 93% in Seattle (Choudhry and Helba, this issue). Excepting Seattle, then, sites using RMA lists included only about two-thirds of the

potentially eligible population of DA&A recipients in their sampling frames.² In contrast, the two sites that used the SSA lists (Los Angeles and Chicago) covered the entire DA&A populations in their catchment areas. The lack of a complete sampling frame at many of the sites led to an ancillary study to determine the representativeness of the sample at the local level, at the aggregate level across sites, and at the national level (Wittenburg et al., this issue). This study showed that the aggregate DA&A sample *was* representative of the aggregate DA&A population across sites but that there were differences at the local and national levels between the samples attained and the relevant populations. Differences between the aggregate DA&A sample and the national population of DA&A recipients are not surprising, since the sample was not constructed to be nationally representative. However, the differences between the local samples and populations are more problematic, as each site's sample *was* intended to represent the local population. Consequently, another study was conducted to develop a set of weights that would, at least partially, correct for the incomplete sampling frames used at some of the sites as well as for differences in site sampling rates (Choudhry and Helba, this issue). All studies in this volume using cross-site data analyze them using the weights developed by Choudhry and Helba.

Sample size determinations

There was considerable variation among the sites with respect to both the targeted and the attained sample sizes. A number of factors drove these differences. First, the size of the DA&A population varied among the sites, from 286 in Portland to 13,997 in Chicago. Second, sites wanted to address different research questions. In Seattle, providing accurate population estimates of various factors and conditions was paramount. There, the priority was to obtain the largest sample possible to permit the greatest precision in providing population estimates. In contrast, the four Northern California sites were interested in hypothesis testing and in making comparisons among the four counties across time using analytic

procedures such as repeated measures analyses. In these counties, sample sizes were driven by estimates of statistical power in order to provide adequate numbers of subjects for such comparative analyses. In Portland, the DA&A population was so small (286 cases) that the research team attempted to recruit the entire population. Third, funding levels varied by site, contingent in part on the presence of an additional funding source such as the RWJ as well as variations in the cost of conducting interviews and recruiting subjects. Fourth, even after setting minimum target sample sizes, there were differences in recruitment rates across the sites (see Choudhry and Helba, this issue) that ultimately determined the sizes of the baseline samples attained.

Finally, another source of sample-size variation among the sites was the differing implementation of the Study Group's general goal for each site to interview at least 200 subjects at baseline. The Study Group set this figure to obtain a sample from each site that would provide local population estimates at the 90% confidence level with a $\pm 5\%$ precision level. The sample-size goal for respondents with complete data after five interview waves was set at 135. For the purpose of estimating a baseline sample size, a 30% cumulative attrition rate was assumed. Based on the projected need to end with completed interviews from 135 subjects, this meant that each site had to interview approximately 200 subjects at baseline. Further, assuming a 70% recruitment rate, it was estimated that each site would have to attempt to contact about 275 subjects in order to have a baseline sample size of 200. The Northern California research team decided that their four counties should *collectively* yield a baseline sample of at least 200 cases (factoring in the power analyses noted above). Seattle sought a sample of 200 subjects who retained SSI benefits and 200 who lost them, which led to that site's having the largest baseline sample of all. And in Portland, as mentioned, the population size was so close to the targeted

sample size that researchers there decided to attempt baseline interviews with all DA&A beneficiaries in that area.

As a result of these factors, baseline sample sizes ranged from 66 subjects in Santa Clara County (San Jose) to 321 subjects in Seattle. The differences in baseline sample sizes, along with some deviations between the obtained local samples and their corresponding population on demographic characteristics, led, in combination, to the development of the sample weights. One aspect of the weights takes into account the variations in recruitment rates and sampling ratios to produce a composite sample that better reflects the DA&A population distributions of the sites.

Survey interview instrument

The Study Group's survey interview instrument was comprised of 582 items and included items from the Global Appraisal of Individual Needs (GAIN) questionnaire (Dennis et al., 1996), questions on hunger derived from the Community Childhood Hunger Identification Project food security questions (Wehler et al., 1996), and composite index items from the Addiction Severity Index (ASI) (McLellan et al., 1992). The ASI composite items were included in seven of the content areas: employment/income, medical, drug use, alcohol use, legal, family/social, and psychiatric status. By adding the ASI composite items, the Study Group intended to provide a standard set of indices for comparing DA&A subjects with other populations of interest such as those entering drug treatment or other groups affected by welfare reform legislation (see Guldish et al., this issue).

The survey instrument, which required about 60 to 90 minutes to administer, included basic demographic questions as well as questions in 10 content areas. As shown in Table 2, the content areas covered a broad spectrum of issues (e.g., drug use, family conflicts, medical and psychiatric problems) and statuses (e.g., medical insurance status, requalification for SSI benefits, employment). Most questions were formatted as yes/no, multiple-choice or fill-in-the-blank items.

However, in a few areas where the Study Group thought a more in-depth response would be informative, the instrument included questions that encouraged verbatim responses in which subjects could elaborate on and clarify their responses to other items on the questionnaire.

The Study Group made changes to the baseline interview instrument over the course of the study. After interviews one, two, and three, the group met to assess the need for instrument revisions. Field reports from some interviewers prompted the reexamination of the instrument by pointing out questions that were ambiguously worded or difficult for respondents to understand. For example, the wording of the question on how many children a respondent “had” caused some respondents to interpret it as meaning how many children were living with them. However, the intent of the question was to determine how many biological, adopted, or foster children the respondent “had” regardless of where the children were living. The Study Group elected to change the wording of this and other problematic questions for subsequent interview periods. The Study Group made changes only if interviewers at more than one site reported that respondents had difficulty with a question or questions. Other changes to the instrument over the course of the study included minor editorial and layout adjustments to make it easier for interviewers to follow the intended flow of the questions.

Over the course of the research, the Study Group also added items to update its understanding of study participants’ status in the appeals process, to alter the time frames of some questions (e.g., to ask about the six months since the last interview instead of the six months between June and December 1996), and to clarify ambiguous or contradictory responses in earlier waves. The largest addition to the instrument, included only for the 12-month follow-up, was a section on respondents’ military service, type of discharge, and use of Veterans Health Administration benefits, and questions were added to the alcohol and other drug section to obtain estimates of life-

TABLE 2
Interview instrument section and content areas and substantive changes across interview periods

Section	Content Area Covered	Substantive Changes/Additions
Demographics	This section captured basic demographic information such as gender, ethnicity, and age.	No substantive changes.
Drug Use (including alcohol use)	ASI items comprised the majority of questions in this section. The questions asked concerned respondents' alcohol and other drug use for the past 30 days and past six months, the amount of money spent on drugs and alcohol, the respondents' ratings of the severity of their substance abuse, and the extent of their need for treatment.	Questions about lifetime drug use for all drugs including alcohol were added at the 12-month follow-up interview only. Questions clarifying methadone use were also added.
Drug Treatment	This section included questions about alcohol and drug treatment utilization by modality and setting (e.g., detoxification, residential, outpatient, hospital, jail/prison, methadone maintenance), treatment payment source(s), perceived need for treatment, and reasons for participating (or not participating) in treatment. It also included questions about self-help group participation.	Questions about use in the previous three days were added at the 12-month follow-up. A series of yes/no response questions asking why the respondent left treatment at baseline and six months were changed to capture only the first three responses at the third interview wave. Questions distinguishing between methadone detoxification and treatment were added at the 12-month follow-up.

Living Arrangements	<p>Questions asked respondents to report on their current living arrangements and whether or not they expected these to change in the near future. These questions were used to assess the proportion of homeless respondents or respondents who were experiencing considerable instability in their housing arrangements.</p>	<p>A question on the kind of quarters for which rent was paid was changed at the six-month follow-up interview to add “rooming or boarding house” and “group home” to baseline’s “someone else’s apartment or house.”</p> <p>Questions asking if the respondent could stay in his/her current residence and if not where he/she planned to stay were dropped after baseline.</p>
Income and Employment	<p>Questions on employment gathered information on legal and illegal sources of income. Respondents reported on their receipt of income from work (legitimate and under the table), SSI, several other forms of social insurance and public assistance, gifts from charitable organizations or family and friends, and illegal activities such as prostitution and selling drugs. Other questions in this section asked respondents about their recent work history and their willingness and perceived ability to work.</p>	<p>At the six-month follow-up, questions were added to ask, “How many people do you live with?” and “Of the people you live with, both adults and children, how many do you share expenses with?”</p> <p>Questions about veteran status were added at the 12-month follow-up interview only.</p> <p>Wording and instructions for the income table were revised and clarified (e.g., separation and clarification of the meaning of “legal” and “illegal” income sources).</p>

TABLE 2 CONTINUES

TABLE 2 CONTINUED

Section	Content Area Covered
Substantive Changes/Additions	<p>Divided a single item ("Have you had enough money for food and transportation?") into two separate items (money for food and money for transportation).</p> <p>Questions using the term "job" were expanded to denote "main job" for those with more than one job at the six-month follow-up and thereafter.</p> <p>At six months, a question was added to ask: "When you received your last SSI check, did it come to you or your representative payee?"</p> <p>Added two questions for the 12-month interview only: "In what ways has receiving SSI been good for you?" and "In what ways has receiving SSI been bad for you?"</p> <p>A series of questions on conflict with family members and friends was changed at 12 months from a Likert scale to yes/no responses.</p>
Family	<p>This section included questions asking respondents for the ages of their children, their children's residence at the time of interview, and any change in those living arrangements over the preceding six months.</p>

Other questions asked about the respondents' marital status and general satisfaction with it, questions from the ASI family/social index about level of contact and conflict with family members, and respondents' perceived need for counseling for these relationships.

Experiences with SSI

This section covered respondents' lifetime and current SSI status, including primary impairments, whether they applied for redetermination or filed an appeal if denied, and if so, on what basis they filed and where their case was in the appeal process. To clarify current SSI status, additional questions asked about the receipt of checks in the months preceding the interview. Other questions asked about retroactive benefit payments ("back pay") for the period from initial application to the time of award. If a lump-sum retroactive payment was received, the respondent was asked to recall how the money was spent. A portion of this section was devoted to questions about services performed by representative payees and respondents' satisfaction with those relationships.

After baseline, a question asking how long the respondent had been receiving SSI benefits was dropped.

Added the following question for the six-month follow-up only: "When did you find out that your SSI benefits for drug or alcohol abuse were to be cut on Jan. 1, 1997?"

At six months a calendar was added to record each month in the past six when a respondent received an SSI check.

Added the following question at six months: "When did you or someone else first give the Social Security Administration the paperwork you thought they needed to qualify you for SSI benefits after the Jan. 1st cutoff?"

TABLE 2 CONTINUES

TABLE 2 CONTINUED

Section	Content Area Covered	Substantive Changes/Additions
Physical Health and Health Care	<p>Questions from the ASI medical index, among other items, asked respondents about their physical health in the past 30 days and whether they had ever been diagnosed as HIV positive. Other questions asked respondents to report on the status and type of their health insurance, their use of health care services in the past six</p>	<p>Verbatim responses to “Have you appealed?” and “How did you find out you needed to appeal?” were coded into a yes/no series of questions after baseline.</p> <p>At six months, the “still pending” response category for SSI redetermination question was split into two responses: “still waiting for a decision” and “additional information requested.”</p> <p>Responses to disability type at baseline were also reformatted after baseline into a series of yes/no items.</p> <p>A question asking why a respondent did not appeal was dropped at 12 months.</p> <p>No substantive changes.</p>

months, and, when appropriate, why they did not seek medical care when they needed it.

Food Sources and Hunger

Questions in this section were derived from the Community Childhood Hunger Identification Project food security questions and included questions on adult hunger and sources of emergency food (see CCHIP, 1995; Scott et al., 1999). Other items assessed income relative to basic expenditures and asked about sources of help with food and housing costs.

A series of questions on “how much it costs to live each month” was added at 12 months.

Law, Crime, and Victimization

The law and crime section incorporated ASI items to examine the range and seriousness of recent criminal activity and arrests. Several questions asked respondents to report on their current legal status and any perceived changes in that status compared with the preceding six months. Additional questions in this section asked how often respondents had themselves been victims of various types of crimes and of physical or sexual abuse.

At 18 months, a question asking “What were you arrested for?” was added and the last four most recent arrests were recorded.

TABLE 2 CONTINUED

Section	Content Area Covered	Substantive Changes/Additions
Mental Health, Distress, and Need for Help	The interview section on distress and mental health incorporated ASI questions and examined respondents' recent experiences with depression, anxiety, and other emotional problems, and their use of psychiatric medications. At 18 months, the question about psychiatric medications was changed to distinguish clearly between medications that were prescribed and those that actually were taken. Interviewers asked respondents to describe their need for help performing necessary tasks (e.g., shopping, cooking, and paying bills) and whether they were getting any help performing these tasks as needed.	Wording of the interviewer instructions was changed at 12 and 18 months to clarify how the questions should be asked if the respondent indicated he/she either did or did not need help with a given activity.
Interviewer Assessment		At 12 months, two locations were added to the list of interview locations ("Survey Research Lab" and "other") and an item indicating whether the respondent provided a urine specimen was added.

time use. The Study Group completed all such changes and additions by the third interview (i.e., at the one-year follow-up) and made no further substantive changes to the questionnaire.

Individual sites could supplement the core instrument to address research issues of particular concern, and seven of the nine participating sites did so; only Portland and Seattle used the core instrument without any supplemental questions or forms at all waves. However, there were variations in the supplements used and in how administration of the supplemental questions was coordinated with administration of the core instrument. The Los Angeles team opted to supplement the instrument with one or more additional forms that addressed different content areas at each wave. The Los Angeles questionnaire thus consisted of the intact core instrument with supplemental forms administered after the core instrument was completed. Research staff in Chicago administered a supplemental instrument collecting psychiatric diagnostic data subsequent to administering the core instrument at the 12-month follow-up only. At the 18-month and 24-month follow-up interviews, Detroit investigators added supplemental questions about contributions respondents made to family and community and to get more details on psychological well-being. As in Los Angeles and Chicago, supplemental questions were administered in Detroit only after the core instrument had been completed. Thus in Los Angeles, Chicago, and Detroit the sequence and context of the core questionnaire items were preserved.

The research team at the four Northern California sites—San Francisco, Stockton, Oakland, and San Jose—took a different approach. They developed a longer instrument that incorporated additional questions within every content area. At these sites, the questionnaire consisted of the core items with embedded questions interspersed throughout. The longer instrument used at these sites altered both the sequence and the context of the core questions. Lacking valid comparison

tools and studies, we cannot estimate the effects on data validity and comparability of changes in the wording of some items across follow-up interviews, or the effects of supplementing the core instrument in different ways and with different questions and forms, most notably in the four Northern California sites.

Urinalysis data After reviewing the baseline and six-month follow-up data, some members of the Study Group were concerned about the relatively low rates of drug use reported by subjects who, as a group, were ostensibly defined by high rates of problematic substance use. Therefore, the Study Group added questions on lifetime drug use for the 12-month follow-up in an attempt to determine if the low rates of self-reported use were the result of substantial underreporting of current use or if the lifetime rates of drug use for this population were simply lower than expected.

As a further response to the low reported rates of drug use, three sites that could gain Institutional Review Board (IRB) clearance opted to collect and test urine specimens for drug use beginning with the 12-month follow-up to determine if the self-reported rates were valid and, if not, to assess the actual rates of use for selected drugs.³ The three sites were Chicago, Los Angeles and Seattle.⁴ To assess the validity of self-reported use and to contribute data to the cross-site database, each of these sites tested urine samples for five substances—marijuana, cocaine, opiates, amphetamines, and PCP—using Enzyme Multiplied Immunoassay Technology (EMIT).⁵ In Chicago the laboratory conducted testing on all five drugs, using the detection thresholds established by the National Institute on Drug Abuse (NIDA).⁶ Seattle followed the NIDA thresholds for all drugs except THC (the active ingredient in marijuana) for which they used a lower detection level, 20 ng/ml (as compared with 50 ng/ml).⁷ Similarly, Los Angeles used the NIDA thresholds for all drugs, except its threshold for THC was higher than the NIDA threshold—100 ng/ml. The detection window for cocaine, opiates, and

amphetamines is for drug use within 24 to 72 hours of collection of the urine sample. For marijuana and PCP the detection window can be much longer, up to two weeks or even a month, depending on the amount and frequency of the drug used.

Each of the three sites that conducted urine testing followed a similar collection protocol. At the beginning of the interview, research staff informed subjects that they would be asked to provide a urine sample for drug testing at the conclusion of the interview and that their decision to provide a urine sample was independent of their decision to participate in the interview. Interviewers offered subjects, who were required to complete a separate informed-consent form, an additional, five-dollar fee to provide the specimen. Collected specimens were then sent to a testing laboratory for analysis, with the test results mailed back to the site for data entry and merging with the questionnaire data.

Recruitment and retention protocols

Research staff at all sites used simple random sampling to select potential subjects from their sampling frames. However, recruitment procedures varied considerably across sites in terms of the wording of the initial contact letters (determined to a great extent by the specific requirements of local IRB), the follow-up procedures used to contact those who did not respond to the initial letters, and when and where the initial interviews were scheduled (see tables 3A and 3B for a summary of intersite differences in subject recruitment protocols).

Each site was responsible for training interview staff on administration of the instrument. The exact procedures used for the trainings, the length of the trainings, and interviewer qualifications varied across sites. Because of the short time available to develop the baseline questionnaire and begin the interviews, baseline interviews were conducted without standardized written instructions to provide detailed information on each question and appropriate responses. A standardized cross-site administration guide—a “Q by Q,” as interview

staff called it—was developed for the six-month interviews and was refined and modified for subsequent waves. Interview staff conducted a majority of the interviews in person with respondents, though a small proportion was done over the telephone (see tables 3A and 3B).⁸

Staff at each site used a variety of techniques to maintain contact with respondents and ensure high follow-up rates, though each site could determine its specific techniques. All sites provided some form of payment to subjects for interviews. Some sites provided laminated cards giving subjects contact information and telling them when their next interview was scheduled. Other sites maintained periodic phone

TABLE 3A

Sample selection and data collection protocols by site				
Site	Sampling Frame List	Baseline Sample Size	Recruitment Rate (%)	Baseline Dates
Chicago	SSA	255	56	December 11, 1996– April 1, 1997
Detroit	RMA	201	71	December 3, 1996– February 8, 1997
Seattle	RMA	321	67	December 3, 1996– January 6, 1997
Portland	RMA	182	71	November 27, 1996– February 4, 1997
Oakland	RMA	151	67	December 14, 1996– February 28, 1997
San Jose	RMA	66	67	December 2, 1996– February 25, 1997
Stockton	RMA	154	68	November 30, 1996– February 28, 1997
San Francisco	RMA	153	68	December 1, 1996– February 7, 1997
Los Angeles	SSA	286	75	December 18, 1996– May 1, 1997 ^a

and mail contact. These efforts resulted in low attrition rates. The aggregate 24-month follow-up rate was 82%, with 1,444 of the 1,764 subjects interviewed at baseline completing all five interviews.

Despite the low attrition, sample representativeness may have been compromised to a small extent by low recruitment rates and loss to follow-up. The relatively low recruitment rates at most of the sites may have biased the sample by including only those subjects who were relatively easy to find, most cooperative, and possibly the least troubled. To the extent that this was the case, the generalizability or external validity of the study may have been compromised.

Respondents who failed to complete one or more interviews differed significantly from those who completed all interviews, as there were disproportionate losses of males, Hispanics, and respondents from Chicago. However, there were no significant differences attributable to attrition in the three outcome measures of drug use. Additionally, as three-quarters of the 390 persons lost at the final wave had completed at least three interviews, many could be included in multivariate analytic procedures such as hierarchical linear modeling. Thus, given the relatively small number of subjects lost to attrition and the use of analytic techniques that further minimize this number, differential attrition was likely a minimal threat to the internal validity of the study.

Data cleaning and the composite data set

Following each interview, interviewers checked responses for completeness and consistency. Interviewer supervisors made a second quality check on the instruments; those that were incomplete or had obvious errors were returned to the interviewer for correction and completion.¹⁰ Data were entered into computerized databases at each site using double data entry strategies to minimize errors. All sites double-entered 100% of the data and checked any forms that yielded discrepancies. Each site then conducted preliminary analyses to assess missing data, identify illogical conjunctions (e.g., used

TABLE 3B

Sample selection and collection characteristics by site

Site	Recruitment Protocol	Stipends	Primary Interview Location
Chicago	Letters sent to prospective subjects a week before the start of scheduled data collection with a toll-free number to call to schedule an interview. Field and phone follow-ups attempted 20 times for non-respondents.	\$40 per interview; \$5 per urine sample	SRL offices at University of Illinois (76%)
Detroit	Initial letters sent by the RMA to eligible beneficiaries and their payees explaining the study and asking them to call a toll-free number to schedule an interview. Respondents who did not reply were located using information obtained from residents of last known addresses, payees, and neighbors. Interviewers were instructed to continue to attempt to contact potential respondents unless they specifically heard from <i>that individual</i> that he or she did not want to participate.	\$40 per interview; \$5 per urine sample	Respondent residence (83%)

Seattle	Letters to potential participants and their payees were sent over the signature of the director of the RMA. The letter provided information about the study and asked the recipient to call a toll-free number if he/she did <i>not</i> wish to be contacted to participate. If no call was made, consent to be contacted was assumed, but only one conversion attempt was allowed for individuals who did not call the toll-free number.	\$20 per interview	Respondent residence (63%)
Portland	Letters sent to all sampled DA&A subjects explaining the project and offering a \$50 gift certificate for initial participation (an additional \$5 certificate was later offered for a urine sample). Interviewers also used telephone directories, the Department of Motor Vehicles, and other information systems to locate eligible respondents. Flyers were posted in public locations to reach clients who were hard to locate. Interviewers were instructed to continue to attempt to contact potential respondents unless they specifically heard from <i>that individual</i> that he or she did not want to participate. If the respondent said that he or she did not wish to participate, contacts were discontinued.	\$50 gift certificate per interview; \$5 gift certificate per urine specimen	Portland Services Agency (78%)

TABLE 3B CONTINUES

TABLE 3B CONTINUED

Site	Recruitment Protocol	Stipends	Primary Interview Location
Oakland	Initial letters sent by the RMA to eligible beneficiaries and their payees explaining the study and asking them to call a toll-free number to schedule an interview. Respondents who did not reply were located using information obtained from residents of last known addresses, payees, and neighbors.	\$40 per interview	Respondent residence (46%)
San Jose	As in Oakland	\$40 per interview	Respondent residence (46%)
Stockton	As in Oakland	\$40 per interview	Respondent residence (46%)
San Francisco	As in Oakland	\$40 per interview	Respondent residence (46%)
Los Angeles	Initial letters sent by DARC explaining the study and asking respondents to participate, and offering a \$25 cash incentive. Unanswered letters were followed up with phone calls, and street location techniques were used for most subjects.	\$25 per interview; \$15 to interview at research offices; \$5 for a urine specimen	Respondent residence (56%)

cocaine in past 30 days but not in the past six months), and to correct any identified errors. Project staff at each site then formatted the data according to the cross-site codebooks, stripped client identifiers from the data, and verified that the data files were transmitted to the cross-site data coordination center at the University of Akron. The data coordination center conducted additional missing-data and logic checks and referred identified problems back to the sites for inquiry and correction. After making the necessary corrections, sites with errors in their datasets returned the data to Akron for a final cleaning and quality check. Following this final check, Akron staff merged all of the individual site datasets to form an aggregate cross-site data file, which they redistributed to project researchers.

During preliminary analyses of the aggregate cross-site data, and in the course of the semistructured interview component of the study (see below), researchers at participating sites found additional problems with data consistency that were not caught by either the preliminary site screening or the data-coordination center screening procedures. They subsequently reported all such issues to the data coordination center, where a record was kept of data-related problems and each problem was investigated. Center staff referred site-specific issues back to the submitting site for further clarification, and conveyed cross-site data problems to the quarterly Study Group meetings for discussion and reconciliation.

Validation studies

As discussed above, individual sites undertook validation studies to determine the accuracy of various subject reports. Excepting the comparison of self-reported drug use and urinalysis data and the interviewer assessment of the reliability of subject responses, researchers restricted their analyses to subjects at their own sites. This was because, typically, the

necessary measures for validating subject responses were available only to researchers at the site conducting the validation study. While this limits their generalizability, these studies provide useful information on the accuracy of responses in a variety of content areas covered by the survey instrument.

Reporting of SSI, GA, and other sources of income

For their site, Study Group members in Seattle evaluated the validity of self-reported receipt of SSI, General Assistance (GA), Temporary Assistance for Needy Families (TANF), and food stamps (Campbell, 1998). An analysis of reported receipt of GA benefits as verified by state records produced a high rate of agreement (Kappa = .77). At the 12- and 18-month interviews, 87% and 89% accuracy was obtained for those reporting benefit receipt. Only 4% of respondents at 12 months and 6% at 18 months reported receiving no benefits when state records indicated otherwise. In fact, more than 85% of respondents at both points in time reported their monthly benefit amount accurately (within \$10). At both 12 and 18 months, concordance with reported and actual food stamp receipt produced a moderate rate of agreement (Kappa = .56). Agreement between reports of TANF benefits and administrative records was in the high end of the moderate range (Kappa = .67). Thus for Seattle subjects there was a moderate rate of agreement on food stamps or other kinds of benefits. One caveat with respect to the TANF results is that the survey question was unclear as to whether it was asking parents about their own receipt of TANF benefits or their children's receipt of them. Children may receive TANF benefits when the adult caretaker does not, though the reverse is not permissible. Thus results may understate TANF receipt in the household and overstate receipt by the adult respondent.

Self-reported drug use

As discussed, Study Group members in Seattle, Los Angeles and Chicago collected urinalysis data on large proportions of their samples at the 12-, 18-, and 24-month follow-ups.¹¹ The results of these analyses were then compared with subjects' self-reported drug use for the three most commonly used drugs: marijuana, cocaine, and opiates. At each interview,

interviewers asked subjects if they had used each of these substances within the preceding six months, the past 30 days, and the past three days. To assess the validity of the self-reported information, we compared responses for the past three days with the urinalysis results, since it is the time interval that most closely approximates the detection window for each of these drugs (Harrison, 1995). Table 4 shows the computed Kappa scores for each site, drug, and interview wave after conducting these comparisons.

TABLE 4

Kappa values for self-reported drug use by site, interview wave, and drug

Interview Wave	Marijuana			Cocaine			Opiates		
	12	18	24	12	18	24	12	18	24
Chicago ^a	0.55	0.61	0.41	0.64	0.58	0.60	0.73	0.49	0.55
Los Angeles ^b	0.49	0.43	0.29	0.50	0.56	0.38	0.49	0.47	0.40
Seattle ^c	0.67	0.60	0.40	0.46	0.48	0.48	0.63	0.68	0.58

^a Ns for Chicago at 12, 18, and 24 months were 212, 204, and 178.

^b Ns for Los Angeles at 12, 18, and 24 months were 213, 205, and 212.

^c Ns for Seattle at 12, 18, and 24 months were 215, 232, and 213.

Inspection of the Kappas reported in Table 4 shows that most of the results (24/27 or 89%) were in the moderate range ($.40 \leq \text{Kappa} \leq .70$), meaning that at most waves and at all three sites there was a fair level of agreement between self-reported use of marijuana, cocaine, and opiates and the urinalysis results for these drugs. However, these results tell only part of the story because they reflect the overall level of agreement that includes not only subjects who used drugs and reported that they used drugs but also subjects who did not use drugs and who reported no use. The concordance rates are much lower when the analyses are restricted to those who tested positive for any drug. For example, among 76 Chicago subjects testing positive for cocaine at the 12-month follow-up interview, 27 (36%) denied using the drug in the preceding three days. Similarly, at the 12-month follow-up in Los Angeles, 31 out of 52 (60%) testing positive for opiates denied recent use. The percentages of those testing positive and

denying use for all three sites and interview waves fell between the ranges of 31% and 69%, which is consistent with figures reported from other validity studies using similar measures (see Hser, 1995; Maisto et al., 1990; McNagny and Parker, 1992). Thus reports of drug use, at least for these three drugs and these three sites, underestimate actual use by 30%–70%, but this is similar to the underreporting found in other surveys with other populations. The implication is that point-prevalence rates based on self-reported drug use in the past three days represent lower bound estimates of actual use (see Podus et al., this issue, for a more detailed consideration of the implications of the urinalysis data and underreporting of drug use for DA&A population prevalence estimates). Past research suggests, though, that the prevalence rates based on use within the past 30 days or six months are likely to be somewhat more accurate (see Harrison, 1995).

Self-reported participation in drug treatment

For each of the five interview waves, the Seattle team examined the validity of self-reported participation in drug treatment in its sample. The findings, shown in Table 5, indicate that respondents more accurately reported some forms of treatment than they did others. Kappas indicating agreement between reported participation in treatment and Washington State records of all publicly funded treatment ranged from an anomalous low Kappa of .17 for detoxification at baseline to a very high Kappa of .91 for methadone treatment at the two-year interview. Most of the other Kappa scores, save those for methadone, ranged from .41 to .56, suggesting that subjects were moderately accurate when reporting their participation in different forms of drug treatment. For methadone treatment (MT), the Kappa scores were uniformly high, ranging from .76 to .91.

Inspection of the cross-tabulations on which the non-MT substance abuse treatment Kappas are based shows that respondents were most likely to “overreport” treatment participation; that is, to report being in substance abuse treatment even though there was no corresponding administrative

TABLE 5

Kappa values for self-reported treatment participation compared with Washington State treatment records by interview wave and treatment modality

Interview Wave	Non-MT Substance Abuse Treatment ^a	Detoxification	Methadone ^b
Baseline (N = 321)	0.46	0.17	0.76
6 months (N = 279)	0.47	0.55	0.83
12 months (N = 297)	0.41	0.49	0.83
18 months (N = 291)	0.56	0.47	0.86
24 months (N = 287)	0.43	0.50	0.91

^a Includes outpatient and residential substance abuse treatment but does not include NA or AA or other forms of self-help.

^b Methadone treatment includes methadone maintenance but not methadone detoxification.

record of their participation. For example, of the 72 subjects who reported being in non-MT substance abuse treatment at baseline, 38% had no administrative record of treatment. In contrast, of 175 subjects who reported no participation in treatment at baseline, 17% were participating in treatment according to administrative data. It is noteworthy that overreporting increased to 50%–60% in the follow-up waves while underreporting decreased to 5% or less. Apparent overreporting of treatment participation may occur because the Washington State database does not reflect treatment in privately funded clinics or in clinics providing federally or charitably funded treatment. This may have suppressed the Kappas generally. Apparent underreporting may be explained in part by the time lag involved in conferring official discharge. When a client drops out of treatment, often a discharge is not entered into the administrative database while the provider seeks to re-engage the client.

The mixed picture of agreement rates between reported and actual participation in treatment may also reflect the difficulty respondents had when trying to classify the kind of treatment they received relative to the options presented on the questionnaire. Evidence from the semistructured interviews shows clearly that in some sites (notably in San Francisco and Stockton, and to a lesser extent in Chicago) many subjects had trouble differentiating between outpatient treatment and participation in self-help groups. It is common practice for outpatient programs to incorporate self-help groups in their daily programming and for self-help groups to be conducted in the same facilities as outpatient treatment. Maintaining cognitive distinctions between these types of programs—distinctions that in many jurisdictions are to some extent bureaucratic and administrative conveniences—may not be important to program participants or anything they have been educated about. In contrast, when the type of treatment was clearly demarcated by distinct programmatic characteristics (such as with methadone treatment), subjects had a high degree of reporting accuracy.¹² Overall, these results from the Seattle group suggest that with the exception of methadone we should interpret with caution subject reports of specific kinds of treatment participation and estimations of participation rates based on these reports. Still, reports of participation in any kind of treatment may be more valid and reliable than reports of participation in one specific kind of treatment.

Item
reliability

Los Angeles team members checked for item-response reliability in the core instrument by comparing with answers on it with answers to similar questions on the addendum administered in Los Angeles. They focused on drug use, drug treatment, mental and physical health, and crime and legal status. Two items asking whether the respondent ever injected drugs produced a near-perfect Kappa of .89, and a comparison of yes/no responses on needle sharing had 75% and 90% agreement rates at baseline and six months. Measures on current

participation in formal treatment (i.e., excluding Alcoholics or Narcotics Anonymous or other self-help groups) produced 96% agreement. Agreement for items inquiring about reasons for entering treatment was generally very good, ranging from 58% to 100%.¹³ Items asking about suicide attempts had 100% correspondence, and agreement for items addressing suicidal thoughts and outpatient psychological treatment was 98% and 86%, respectively. Correspondence for reports of inpatient treatment was slightly lower (71%). Comparison of baseline questionnaire items asking participants if they had been in outpatient mental health treatment within the past six months and similar questions on the Los Angeles Addendum yielded a Kappa of .37, with a somewhat higher rate of agreement achieved comparing responses with questions asking about receipt of inpatient mental health treatment (Kappa = .55).

Agreement for responses to questions on physical health, emergency room use, and insurance was quite high, ranging from 76% to 100%. Two items reporting HIV status produced Kappas of 1.0 and .91 at baseline and follow-up. Finally, 100% of those reporting time in jail or prison in the past month also reported spending time in jail or prison in the past 12 months.

The Northern California group conducted an analysis of response coherence¹⁴ on baseline data to determine whether there were differences in reliability of responses in specific domains within the questionnaire. The results showed strong correlations for questions within the categories of reported SSI status (Pearson's $r = .94$), legal work and mental health (.95), under-the-table or off-book work, criminal activity, and physical health (.90), and criminal justice system involvement (.84). High correlations also were found for items on the purchase and use of cigarettes (.98), alcohol (.88), and both alcohol and other drugs (.78). Lower coherence scores were produced by questions on the purchase and use of drugs (.69), household size (.54), and reported income versus

expenses (.16). Thus, for most areas of the questionnaire, respondents answered the questions in ways that were internally consistent.

Interviewers' evaluation of respondent response reliability

The University of Akron research team examined two items asked at each interview that assessed the truthfulness and reliability of information given by the respondent. Interviewers at all five waves reported that most respondents seemed to answer the interview questions honestly. At baseline, 87% of the respondents were evaluated as always or mostly truthful in their responses; the percentage increased for the second wave, to 95%, and the last three waves each produced a 96% rating.

A second question asked interviewers to evaluate the reliability of information given by the respondent. At baseline, 96% of the respondents were judged to be moderately reliable or reliable in the information provided. This proportion dropped at wave 2 (15% of the respondents had missing data for this response at wave 2, but response completion increased for the final three waves to less than 1% missing). The proportions of respondents whose information was evaluated as reliable were 97%, 98%, and 99% for waves 3, 4, and 5, respectively.

Qualitative data collection

After pretesting in Detroit and San Francisco (using an auxiliary sample) during the fall of 1997, a small team of researchers conducted semistructured interviews with 163 of the study's respondents between the middle of March and late May of 1998 (between the 12-month and 18-month survey periods). These interviews were conducted in Portland (n=40), San Francisco (n=40), Stockton (n=41), and Chicago (n=42). Based on their sources of income at wave 3, respondents were randomly selected within three strata: (1) still collecting SSI; (2) not collecting SSI but doing some work for wages; and (3) not collecting SSI and not working for wages.

Forty interviews were scheduled at each site over the course of a week: 13 for each of the three groups of respondents noted above, with the odd interview drawn from the working group in sites where enough respondents reported earnings (Portland, Stockton and Chicago) and from the non-working, no-SSI group where this was not possible (San Francisco). In Stockton and Chicago, circumstances in the field permitted more than 40 interviews, so additional respondents with earnings were chosen from a replacement pool selected in advance. When replacements for original sample members were needed, they were drawn from the same stratum as the dropout. When that was not possible, the first replacement choice was a working person, the second was a respondent who was neither working nor on SSI, and the last resort was a respondent still collecting SSI. Personnel from the local research teams helped respondents keep their appointments, notified replacements when necessary, and in a few cases occupied the children of respondents who brought them along. Except in Portland, where subjects received a \$50 gift certificate, respondents were paid \$40 for the interview.

The interviews were conducted in private offices in social service agencies (in San Francisco and Stockton), a county agency (Portland), and at the offices of the Survey Research Laboratory in Chicago. They were organized to approximate short life histories but structured so that respondents talked about their lives before SSI, while on SSI, and, when appropriate, after SSI. The interviews were designed to supplement the structured interviews in several ways. First, they explored the respondents' experiences with SSI in great detail. Respondents spoke at length about their experiences with representative payees and mandated treatment, their appeals, how they came to apply for SSI, and why they were categorized as DA&A cases. Second, these interviews gathered detailed mental health and substance abuse treatment histories (including participation in self-help groups) and sounded respondents' opinions about treatment and SSI. Third, the

qualitative data provided a much more complete picture of respondents' historical and contemporary relationships with family and friends than is available from the structured-interview data. Similarly, the semistructured interviews captured details of living arrangements and work histories that the structured interviews could not. Finally, because respondents talked about the whole of their lives, the semistructured interviews permitted an appreciation of how a usually brief period of SSI receipt fit into a much longer span of time and with what meaning.

Each team member prepared for an interview by creating an integrated summary of the respondent's first three waves of data, noting ambiguities and inconsistencies. These were explored in the context of the interviews, which lasted from 45 minutes to three hours, with the median length falling between 60 and 90 minutes. As a result, these interviews provided a limited validity check on responses to the structured interviews. However, like the structured interviews, they relied perforce on what people said rather than on observation of what they did. Perhaps most important, the semistructured interviews yielded significant clues and caveats for interpretation of the structured-interview data, as some of the papers in this issue attest.

During the first set of semistructured interviews (in Portland), it became clear that concurrent SSI and DI beneficiaries remained in the study sample despite efforts to eliminate them. This led to the systematic cross-checking of all samples with SSA data. On this basis, seven semistructured-interview respondents (four in Portland and three in Chicago) were ultimately removed from the qualitative database, leaving 156.¹⁵

Of the 156 interviews in the final database, 150 were taped and the others were captured by interviewer's notes. The tapes were transcribed and audited for accuracy; electronic transcripts and notes from the six unrecorded interviews were then entered for analysis into QSR NUD.IST, version 4 (N4).

Conclusions

The SSI study represents an ambitious attempt to conduct a natural field experiment to evaluate the effects of sweeping federal legislation. The study was largely successful in creating a collaboration among a group of researchers from nine sites across the country and in setting up a complex longitudinal survey study that collected data on close to 2,000 subjects across two years and five interview waves. Although the population from which we selected our study sample was not representative of all DA&A recipients nationally, it represented about one-fourth of the national population, was demographically diverse, and resided in settings with diverse arrays of social services and safety nets. Some important flaws notwithstanding, the survey questionnaire that the SSI Study Group developed helped the research team gather a broad array of information about the people most likely to be affected by the legislative changes. The research team was also extremely successful at conducting follow-up interviews and, despite some dropoff at 24 months, achieved a very low attrition rate with a population that was often difficult to track. The SSI Study also successfully included a qualitative component to examine the lives of a subsample to help clarify important issues that could not be addressed within the more structured protocol and format of the longitudinal survey.

Important methodological limitations in the survey study should be acknowledged. The most serious issues are those that potentially affected the internal validity of the study. Some of these derived from the use of a questionnaire that was not validated prior to the start of the study and by continuing change to the instrument. The legislation that ended the DA&A benefit category gave rise to a unique set of circumstances: A relatively large group of presumably heavy or once-heavy alcohol and other drug users, assessed as being disabled by their use, would abruptly lose all federal cash

benefits at the same time and with short notice; many would also lose Medicaid benefits as well. Since there were no existing questionnaires that assessed things like experiences with the Social Security Administration or that completely incorporated the broad number of areas the Study Group wished to examine, the group had little choice but to develop its own questionnaire despite the known methodological risks and the extremely short time in which to do it.

Other internal validity issues have to do with the collaborative nature of the study and variations among the sites in instrumentation and protocols. Because the Study Group discovered a number of problems in the baseline instrument, it made changes to a number of questions and continued to make changes through the first 12 months of follow-up. Changes in the survey instrument over time, however well intended, made it difficult to assess changes over time on the particular issues covered by these questions. More serious in this regard is that the four Northern California sites used an instrument very different from the other sites'. This introduces a potential confound whereby if subjects at these four sites tend to give different responses or show different patterns of behavior over time, we cannot know the extent to which the differences among the sites are attributable to questionnaire differences or to real differences in subject behavior.

Similarly, some sites collected urine specimens, others did not, and some sites collected urine specimens on some subjects some of the time. How this may have affected subjects' responses to the questions on drug use cannot be determined, though we do know from the urinalysis data collected at some sites that the levels of self-reported drug use are most probably lower bound estimates (see Podus, Chang et al., this issue). However, with respect to the validity of the self-reported drug use information, the SSI Study appears to have yielded results similar to those of other surveys of drug use (Hser, 1995).

As demonstrated by Wittenburg et al. (this issue), the sites collected samples that demographically closely resembled the sampling frames from which they were drawn. However, the generalizability of the study was potentially compromised by the use of RMA lists at seven of the nine sites. These excluded some DA&A recipients who otherwise should have been eligible for recruitment into the study. Wittenburg et al.'s analysis mitigates this concern to some extent, and the weights developed by Choudhry and Helba (this issue) provide for an analytic sample that, at least demographically, closely approximates the aggregate population of DA&A beneficiaries represented by the nine participating sites.

In sum, within the context of these limitations, the SSI Study provides useful information on one population affected by the welfare-reform legislation that marked the latter part of the 1990s. Given the limitations just noted, though, readers should take care in making generalizations on the basis of these data, and it should be understood that because of the design issues described in detail above, any estimates derived from the data probably have larger standard errors than those assumed in standard statistical analyses. For this reason, for analyzing the data collected for this study we recommend use of statistical software tools like Wesvar or Sudaan that provide estimates of standard errors to control for design effects.

- Notes**
1. A second, small sample of subjects from San Francisco was added to permit a validation study of the instrument.
 2. In Seattle, just prior to March 1996, the Washington State RMA received from the SSA a data file listing all SSI DA&A recipients who, according to SSA records, were currently assigned to the RMA for monitoring. The Seattle sample was drawn from this list, augmented by the existing RMA list, which included DA&A beneficiaries who might not have been monitored (because they had completed treatment, for example) but remained in active pay status and therefore on the SSI rolls. The sampling frame included only subjects in pay status in October 1996.

3. Los Angeles was the only site to collect and test urine specimens at all five waves.
4. Portland used the collection of urine specimens to conduct a “study within a study” to determine if when the urine sample is collected—before or after the interview—affects the levels of self-reported drug use. Consequently, half of their subjects were randomly assigned to have urines collected before the interview and the other half had urines collected after the interview. These assignments were then reversed at a subsequent interview wave.
5. Chicago and Los Angeles also tested the urine samples for methadone, but this was not reported to the cross-site database. Chicago also conducted confirmation testing on specimens with positive opiates results to discriminate between morphine and codeine derivatives. The results of this testing showed that, for the Chicago sample at least, over 90% of the positive opiate results were attributable to a morphine derivative, most likely heroin.
6. The NIDA thresholds by substance are as follows: amphetamine—1,000 ng/ml; cocaine—300 ng/ml; opiates—300 ng/ml; PCP—25 ng/ml; THC—50 ng/ml.
7. The 20 ng/ml threshold for marijuana detection is low relative to the current NIDA standard of 50 ng/ml. The laboratory used by the Seattle site indicated that they use proprietary tests for detection and confirmation. They describe their initial procedure as an immunoassay test equivalent to EMIT followed by a confirmation test that uses a modified variant of thin layer chromatography (TLC) designed to have higher sensitivity and specificity than standard TLC (Comprehensive Toxicology Services, personal communication).
8. The proportion of interviews conducted over the phone at baseline was 1.6%. At each subsequent wave this proportion increased slightly (2.7% at wave two, 3.5% at wave three, 4.5% at wave four, and 6.3% at wave five).
9. Although the baseline interview times for Chicago lasted through May 1997, the large majority of subjects in that site, over 95%, had been interviewed by the middle of April 1997. Interviews continued past this point to try to meet sample-size goals.
10. The Chicago site used paper-and-pencil interviewing materials only at baseline and the six-month follow-up. At subsequent waves, staff entered questionnaire data directly via laptop computer.
11. Los Angeles began collecting urinalysis data at baseline, but because these data were collected using money from other funders, technically they were not part of the CSAT-sponsored SSI Study. Because of the large sample size, Seattle randomly selected about half of their

subjects for urine collection. The Portland and Detroit sites also collected urinalysis data beginning with the 12-month interview, but they used the data to support an experimental study to see if the collection of urine specimens and the timing of when subjects were notified of the collection of urine samples would affect drug-use response rates. Further, only a small number of subjects participated at the Detroit site, and for only one collection period. Because of the potential confounds introduced by their experimental manipulations, the relatively small number of subjects involved, and the complexity of analyzing these data in conjunction with the urinalysis data collected at the three other sites, we did not include data from these two sites in the analyses presented here.

12. Even so, Washington State administrative data and the semistructured interviews make clear that clients often report their methadone-program-related outpatient counseling as a separate form of treatment.
13. The core instrument and Los Angeles addendum questions were not fully parallel; hence percentages instead of Kappas are reported for some results in this section. For example, the core instrument asked for yes/no responses to a series of questions: Were you in treatment because SSI required it? (Y/N); Were you in treatment because you were required to attend by the criminal justice system? (Y/N); Were you in treatment because of health concerns? (Y/N); Were you in treatment because of pressure from family and friends? (Y/N). By contrast, the Los Angeles addendum question was open-ended. People were asked to state the most important reasons why they attended treatment—up to a maximum of two responses. Because the addendum item was open-ended and limited to a maximum of two responses, failure to mention an item asked on the core instrument list did not mean that it was not a reason, but that it was not one of the top two reasons. Because the format was different, the actual comparison made was between those who gave a certain answer when asked the open-ended addendum item (e.g., the SSI mandate was one of the two most important reasons they were in treatment) and those who responded affirmatively when asked directly as part of the core instrument list whether that was a reason (e.g., Did you attend treatment because of the SSI mandate? Y/N).
14. While the Kappa statistic measures complete agreement between items, coherence is a measure of the average level of consistency for all respondents within a particular grouping of responses. Individuals were assigned a score of zero or one for each item within an area (zero for non-agreement, one for agreement). If one or more items within an area had a value of one, the respondent was assigned a one for agreement in the area. Mean overall coherence scores (for each area and all areas combined) were calculated for each respondent and

compared using the Pearson's correlation coefficient to the group means.

15. Even after this cross-check, one case of reported concurrent benefits remains in the Stockton semistructured interview group. As this subject had no motive to misrepresent his benefit status, and as he reported receiving two checks of different colors on different days of the month—as concurrent beneficiaries would—we suspect that SSA administrative files are not an infallible standard for distinguishing SSI-only beneficiaries. There may be a few concurrent beneficiaries remaining in the various site samples, and it is also possible that the study missed a small number of SSI-only recipients miscoded by the SSA as concurrent beneficiaries.

References

- Campbell, Kevin (March 1998). *Anticipating the Impact of Eliminating the SSI Drug Addiction and Alcoholism Category: Validation of Self-Reported SSI Benefit Status, Participation in Substance Abuse Treatment, and Participation in State-Funded Social Service Benefits Programs*. Paper presented to the SSI Research Group.
- Choudhry, G. Hussain and Cynthia Helba (2003), "Rationale and Procedures for Weighting the SSI Study Data," *Contemporary Drug Problems*, 30: 137-146.
- Community Childhood Hunger Identification Project [CCHIP] (July 1995). *A Survey of Childhood Hunger in the United States*. Food Research and Action Center, Washington, DC.
- Dennis, M.L., Webber, J.R., White, W.L., Senay, E., Adams, L., Bokos, P., Eisenberg, S., Fraser, J., Moran, M., Ravine, E., Rosenfeld, J. & Sodetz, A. (1996). *Global Appraisal of Individual Needs (GAIN), Vol. 1: Administration, Scoring and Interpretation*. Bloomington, IL: Chestnut Health Systems.
- Guydish, Joseph, Claudia Ponath, Alan Bostrom, Kevin Campbell, and Nancy Barron (2003). "Effects of Losing SSI Benefits on Standard Drug and Alcohol Outcome Measures," *Contemporary Drug Problems*, 30: 169-193.
- Harrison, L. (1995). The validity of self-reported drug use in survey research: An overview and critique of research methods. In L. Harrison & A. Hughes (eds.), *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*. NIDA Research Monograph 167 (pp. 17-36). Rockville, MD: National Institute on Drug Abuse.

- Hser, Y. (1995). Self-reported drug use: Results of selected empirical investigations of validity. In L. Harrison & A. Hughes (eds.), *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates. NIDA Research Monograph 167* (pp. 320-343). Rockville, MD: National Institute on Drug Abuse.
- Hunt, Sharon R. and Jim Baumohl (2003). "Drink, Drugs and Disability: An Introduction to the Controversy," *Contemporary Drug Problems*, 30: 9-76.
- Maisto, S.A., McKay, J.R. & Connors, G.J. (1990). Self-report issues in substance abuse: State of the art and future directions. *Behavioral Assessment*, 12: 117-134.
- McLellan, A.T., Kushner, H., Metzger, D., Peters, R. et al. (1992). The fifth edition of the addiction severity index. *Journal of Substance Abuse Treatment*, 9(3): 199-213.
- McNagny, S.E. and Parker, R.M. (1992). High prevalence of recent cocaine use and the unreliability of patient self-report in an inner-city walk-in clinic. *Journal of the American Medical Association*, 267(8): 1106-1108.
- Podus, Deborah, Eunice Chang, Mary Lynn Brecht, James Swartz, and M. Douglas Anglin (2003). "Drug Use Prevalence Among Former SSI DA&A Recipients," *Contemporary Drug Problems*, 30: 275-290.
- Scott, Richard Ira, Wehler, Cheryl A., Anderson, Jennifer J. (November 1999). Measurement of Coping Behaviors as an Aspect of Food Insecurity. Washington, DC: Food Research and Action Center. Presented at Food Security Measurement and Research Conference sponsored by the Food and Nutrition Service (USDA), the National Center for Health Statistics (NIH), January 1994, Washington, DC.
- Wehler, C.A., Scott, R.I., Anderson, J.J. (1996). Development and testing process of the Community Childhood Hunger Identification Project Scaled Hunger Measure and its application for a general population survey. In: *Conference on Food Security Measurement and Research: Papers and Proceedings*, January 1994, Washington, DC, Technical Appendix A. Alexandria, VA: United States Department of Agriculture, Food and Consumer Service.
- Wittenburg, David, David Stapleton, Adam Tucker, and Rick Harwood (2003). "An Assessment of the Representativeness of the SSI DA&A Study Panels," *Contemporary Drug Problems*, 30: 123-136.