SAMPLE DESIGN, SELECTION AND ESTIMATION FOR PHASE I OF ADSS

Final Report

Prepared for:

Substance Abuse and
Mental Health Services
Administration

Prepared by:

Leyla Mohadjer
Ibrahim Yansaneh
Thomas Krenzke
Sylvia Dohrmann

Westat
1650 Research Boulevard
Rockville, Maryland 20850

January 2000

ACKNOWLEDGMENTS

This report was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies (OAS), by Westat Inc., and the Schneider Institute for Health Policy, Brandeis University, under Contract No. 283-92-8331. Constance M. Horgan was the Principal Investigator. Helen J. Levine was the Co-Principal Investigator and project director for the ADSS Phase I Facility Survey. Westat, Inc., subcontractor on the project, conducted the data collection. Anita Gadzuk was the SAMHSA Project Officer and the lead OAS reviewer for this report.

Authors: Leyla Mohadjer, Ibrahim Yansaneh, Thomas Krenzke, and Sylvia Dohrmann, all of Westat.

PUBLIC DOMAIN NOTICE

COPIES OF THE REPORT

Copies of this report may be obtained, free of charge, by calling the Office of Applied Studies, SAMHSA at (301) 443-6239.

ORIGINATING OFFICE

SAMHSA, Office of Applied Studies
5600 Fishers Lane, Room 16-105
Rockville, MD 20857

January 2000

SUBSTANCE ABUSE AND MENTAL HEALTH SERVICES ADMINISTRATION
U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

List of Appendices

List of Tables

List of Tables (continued)

TABLE OF CONTENTS (Continued)

List of Tables (continued)

TABLE OF CONTENTS (Continued)

List of Tables (continued)

List of Exhibit

1. INTRODUCTION AND OVERVIEW

The sample for the 1996 ADSS survey was selected using a multi-stage stratified design, with selection of facilities as the first stage (Phase I), and selection of a subset of Phase I responding facilities within 62 predetermined PSUs and selection of client discharge records as the second stage (Phase II). Clients selected as part of the second stage will be followed up with interviews in Phase III of the survey. However, the methadone clients in the discharge sample selected in Phase II will not be followed up in Phase III. This report discusses the sample design, selection and estimation processes for Phase I of the survey.

The Phase I sample is a stratified sample of 2,395 responding facilities in the seven sampling strata as defined later in this section. The targeted sample size of respondents was 2,140 as shown in Table 1-1. An analysis of the 1990 Drug Services Research Survey (DSRS) report showed that a minimum of about 300 facilities per stratum is required to assure reasonably precise and stable estimates of statistics of interest. In some cases, a stratum may be larger than 300, as is the case for stratum 5, "All Other Outpatient", which is set at about 500 to account for the larger population in this stratum (see Table 1-1). Stratum 7 includes facilities for which no information on treatment modality and number of clients is available. As indicated later in this report, about 19 percent of the facilities in the frame are in stratum 7. To derive the Phase I sampling rates, and draw the sample, a number of assumptions were made about the composition of facilities (by treatment modality) in stratum 7. Based on these assumptions, a measure of size was assigned to each of the facilities with unknown number of clients, prior to sample selection.

At about the same time as the ADSS survey, another national survey was conducted to evaluate substance abuse treatment programs. Westat has a contract with the Tennessee Valley Authority (TVA) to design and conduct the National Evaluation of Substance Abuse Treatment (NESAT) with the Center on Addiction and Substance Abuse (CASA) at Columbia University, for the Office of National Drug Control Policy (ONDCP). As agreed with SAMHSA, the sample selection design for ADSS was revised to minimize the overlap between the ADSS and NESAT surveys to reduce the respondent burden that will be imposed on substance abuse treatment facilities selected in both surveys, and thereby increase the response rate. A description of the approach used to minimize overlap is given in Section 3.6.1.

Table 1-1. Counts of facilities and clients in treatment, by stratum for the entire ADSS sampling frame

Stratum[1]	Facilities		Clients	Average number of clients per facility	Phase I target sample sizes
	Total	Percent	Total		
Hospital Inpatient	1,168	6.4	12,255	10	316
Other Residential	2,329	12.7	73,280	31	316
Outpatient – PM[2]	511	2.8	111,047	217	315
Outpatient – AEA[3]	2,063	11.2	201,830	98	316
Outpatient – AO[4]	6,224	33.9	523,347	84	560
Combined	2,575	14.0	255,550	99	317
Unknown	3,498	19.0	NA	NA	--
Total	18,368	100.0	NA	64	2,140

[1] Refer to Section 3-1 for details of stratification.

[2] PM = Predominantly Methadone.

[3] AEA = Almost Exclusively Alcohol.

[4] AO = All Other.

This report provides details of the procedures employed in the design, selection, weighting and imputation of the sample of facilities for Phase I of ADSS. Section 2 describes the construction, enhancement, and preparation of the ADSS sampling frame prior to sample selection. The distributions of facilities and clients by state, sampling strata, and other sampling variables are also presented. A detailed description of the sample selection procedure is provided in Section 3. The assignment of a measure of size to each facility on the sampling frame prior to sample selection is discussed. The issue of overlap between the ADSS and NESAT surveys is also addressed. The process of selection and unduplication of the sample of facilities for ADSS Phase I is also described in detail. The assignment of the appropriate probabilities of selection to the unduplicated sample is discussed in Section 3.11.3. Other issues discussed in Section 3 include administrative units, the computation of preliminary base weights, and stratum migration.

Finally, Sections 4 and 5 describe the estimation methodology. Section 4 describes the process of generating full sample and replicate weights that account for unit nonresponse in the sample. The procedures used to compute the base weights is explained in Section 4.1.3. Several adjustments to the weights were implemented, and the reasons and procedures are explained in the remainder of Section 4.1.3. The process of generating replicate weights is discussed in Section 4.1.4. Section 5, specifically Sections 5.4 and 5.5, discuss imputation procedures that were used to account for item nonresponse. Section 5.10 discusses the impact that the imputed values may have on the survey estimates. Section 5.11 explains how to account for the imputation error variance when analyzing data.

2. THE ADSS SAMPLING FRAME

The ADSS sampling frame was constructed with the objective of covering all substance abuse treatment facilities that have active treatment programs in all fifty states and the District of Columbia. The sampling frame consists of public and private substance abuse treatment facilities.

The frame consists of two major components: active facilities offering substance abuse treatment programs as listed in SAMHSA's National Facility Register (NFR) as of September 1995, and the enhancement file. More than three-quarters of the facilities to be included in the ADSS sampling frame come from the NFR file and the remainder come from the enhancement file (see Table 2-1).

2.1 Exclusions From the ADSS Frame

Treatment facilities of the following types are excluded from consideration for ADSS:

- Halfway Houses with no paid treatment staff;
- Solo Practitioners;
- Jails/Prisons;
- Military/DoD;
- Indian Health Service; and
- Intake and Referral only.

Facilities known to be ineligible for ADSS, for instance, facilities operated by the Bureau of Prisons (BOP), the Department of Defense (DoD), and the Indian Health Service (IHS) were dropped from the ADSS sampling frame using the associated information in the frame, and the rest were designated as ineligible during the screening of sampled facilities in Phase I.

Table 2-1. Distribution of facilities with known or unknown number of clients in the NFR, and the Enhancement File by state

	NFR						Enhancement file						Total					
	Known		Unknown		Total		Known		Unknown		Total		Known		Unknown		Total	
State	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
AK	45	54.9	27	32.9	72	87.8	8	9.8	2	2.4	10	12.2	53	64.6	29	35.4	82	100.0
AL	75	54.3	4	2.9	79	57.2	59	42.8	0	0.0	59	42.8	134	97.1	4	2.9	138	100.0
AR	53	55.8	9	9.5	62	65.3	23	24.2	10	10.5	33	34.7	76	80.0	19	20.0	95	100.0
AZ	141	57.8	17	7.0	158	64.8	74	30.3	12	4.9	86	35.2	215	88.1	29	11.9	244	100.0
CA	1,379	66.0	269	12.9	1,648	78.9	394	18.9	48	2.3	442	21.1	1,773	84.8	317	15.2	2,090	100.0
CO	156	57.6	26	9.6	182	67.2	77	28.4	12	4.4	89	32.8	233	86.0	38	14.0	271	100.0
CT	201	62.8	22	6.9	223	69.7	80	25.0	17	5.3	97	30.3	281	87.8	39	12.2	320	100.0
DC	55	61.8	22	24.7	77	86.5	10	11.2	2	2.2	12	13.5	65	73.0	24	27.0	89	100.0
DE	45	60.0	18	24.0	63	84.0	7	9.3	5	6.7	12	16.0	52	69.3	23	30.7	75	100.0
FL	623	47.1	527	39.9	1,150	87.0	144	10.9	28	2.1	172	13.0	767	58.0	555	42.0	1,322	100.0
GA	108	37.2	8	2.8	116	40.0	152	52.4	22	7.6	174	60.0	260	89.7	30	10.3	290	100.0
HI	53	60.2	14	15.9	67	76.1	19	21.6	2	2.3	21	23.9	72	81.8	16	18.2	88	100.0
IA	73	48.0	14	9.2	87	57.2	48	31.6	17	11.2	65	42.8	121	79.6	31	20.4	152	100.0
ID	22	27.2	15	18.5	37	45.7	38	46.9	6	7.4	44	54.3	60	74.1	21	25.9	81	100.0
IL	394	51.6	250	32.8	644	84.4	88	11.5	31	4.1	119	15.6	482	63.2	281	36.8	763	100.0
IN	150	45.2	38	11.4	188	56.6	113	34.0	31	9.3	144	43.4	263	79.2	69	20.8	332	100.0
KS	165	56.1	64	21.8	229	77.9	51	17.3	14	4.8	65	22.1	216	73.5	78	26.5	294	100.0
KY	254	59.9	105	24.8	359	84.7	51	12.0	14	3.3	65	15.3	305	71.9	119	28.1	424	100.0
LA	126	54.5	15	6.5	141	61.0	75	32.5	15	6.5	90	39.0	201	87.0	30	13.0	231	100.0
MA	277	63.4	6	1.4	283	64.8	138	31.6	16	3.7	154	35.2	415	95.0	22	5.0	437	100.0
MD	293	69.4	56	13.3	349	82.7	64	15.2	9	2.1	73	17.3	357	84.6	65	15.4	422	100.0
ME	140	67.6	35	16.9	175	84.5	26	12.6	6	2.9	32	15.5	166	80.2	41	19.8	207	100.0
MI	590	76.4	114	14.8	704	91.2	53	6.9	15	1.9	68	8.8	643	83.3	129	16.7	772	100.0
MN	258	73.9	41	11.7	299	85.7	42	12.0	8	2.3	50	14.3	300	86.0	49	14.0	349	100.0
MO	139	52.1	18	6.7	157	58.8	94	35.2	16	6.0	110	41.2	233	87.3	34	12.7	267	100.0
MS	65	45.8	0	0.0	65	45.8	64	45.1	13	9.2	77	54.2	129	90.8	13	9.2	142	100.0
MT	27	39.1	3	4.3	30	43.5	35	50.7	4	5.8	39	56.5	62	89.9	7	10.1	69	100.0
NC	125	38.3	5	1.5	130	39.9	161	49.4	35	10.7	196	60.1	286	87.7	40	12.3	326	100.0
ND	43	71.7	10	16.7	53	88.3	6	10.0	1	1.7	7	11.7	49	81.7	11	18.3	60	100.0
NE	114	65.1	22	12.6	136	77.7	34	19.4	5	2.9	39	22.3	148	84.6	27	15.4	175	100.0

Table 2-1. Distribution of facilities with known or unknown number of clients in the NFR, and the Enhancement File by state (continued)

	NFR						Enhancement file						Total					
	Known		Unknown		Total		Known		Unknown		Total		Known		Unknown		Total	
State	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
NH	48	50.0	13	13.5	61	63.5	30	31.3	5	5.2	35	36.5	78	81.3	18	18.8	96	100.0
NJ	282	60.9	16	3.5	298	64.4	148	32.0	17	3.7	165	35.6	430	92.9	33	7.1	463	100.0
NM	63	50.8	3	2.4	66	53.2	47	37.9	11	8.9	58	46.8	110	88.7	14	11.3	124	100.0
NV	42	48.3	18	20.7	60	69.0	19	21.8	8	9.2	27	31.0	61	70.1	26	29.9	87	100.0
NY	1,209	74.8	158	9.8	1,367	84.5	221	13.7	29	1.8	250	15.5	1,430	88.4	187	11.6	1,617	100.0
OH	438	59.8	154	21.0	592	80.9	109	14.9	31	4.2	140	19.1	547	74.7	185	25.3	732	100.0
OK	93	49.7	20	10.7	113	60.4	59	31.6	15	8.0	74	39.6	152	81.3	35	18.7	187	100.0
OR	149	58.7	37	14.6	186	73.2	61	24.0	7	2.8	68	26.8	210	82.7	44	17.3	254	100.0
PA	566	68.9	155	18.9	721	87.8	91	11.1	9	1.1	100	12.2	657	80.0	164	20.0	821	100.0
RI	70	69.3	13	12.9	83	82.2	13	12.9	5	5.0	18	17.8	83	82.2	18	17.8	101	100.0
SC	74	48.4	25	16.3	99	64.7	41	26.8	13	8.5	54	35.3	115	75.2	38	24.8	153	100.0
SD	47	65.3	1	1.4	48	66.7	21	29.2	3	4.2	24	33.3	68	94.4	4	5.6	72	100.0
TN	87	36.0	4	1.7	91	37.6	122	50.4	29	12.0	151	62.4	209	86.4	33	13.6	242	100.0
TX	572	48.4	318	26.9	890	75.3	239	20.2	53	4.5	292	24.7	811	68.6	371	31.4	1,182	100.0
UT	58	26.6	136	62.4	194	89.0	20	9.2	4	1.8	24	11.0	78	35.8	140	64.2	218	100.0
VA	155	49.7	16	5.1	171	54.8	117	37.5	24	7.7	141	45.2	272	87.2	40	12.8	312	100.0
VT	20	47.6	0	0.0	20	47.6	19	45.2	3	7.1	22	52.4	39	92.9	3	7.1	42	100.0
WA	288	65.0	86	19.4	374	84.4	54	12.2	15	3.4	69	15.6	342	77.2	101	22.8	443	100.0
WI	289	62.7	7	1.5	296	64.2	131	28.4	34	7.4	165	35.8	420	91.1	41	8.9	461	100.0
WV	46	54.1	5	5.9	51	60.0	29	34.1	5	5.9	34	40.0	75	88.2	10	11.8	85	100.0
WY	42	60.9	1	1.4	43	62.3	23	33.3	3	4.3	26	37.7	65	94.2	4	5.8	69	100.0
Total	10,827		2,960		13,787		3,842		739		4,581		14,669		3,699		18,368	

2.2 Frame Enhancement

In an attempt to improve coverage of the ADSS sampling frame, considerable effort was expended on a frame enhancement process prior to the selection of Phase I facilities. Westat was responsible for enhancing that part of the sampling frame that exists within the boundaries of Westat's 62 PSUs (described in Appendix A), and other contractors were responsible for enhancing the parts of the frame that are outside Westat's 62 PSUs.

The frame enhancement process for Westat's 62 PSU portion of the ADSS sampling frame involved comparing NFR records with facility records from supplementary independent sources in a multi-step matching process, and then conducting a telephone screener interview with those facilities not already on the NFR file. Frame enhancement on the parts of the frame that are outside the boundaries of Westat's 62 PSUs was conducted using similar matching and screening procedures as those used by Westat.

The results of the frame enhancement screening, conducted by Westat and the other contractor, were transmitted through SAMHSA to a third contractor who combined the files and produced a final version of the enhancement file for the entire country. This file was then transmitted back to Westat for use on ADSS. The Enhancement File is the list of additional substance abuse treatment facilities using frame enhancement screener records from Westat and other contractors.

2.3 Preparation of the ADSS Sampling Frame for Sample Selection

2.3.1 Duplicate Facilities

A search for duplicate records was conducted on the 14,146 records of active substance abuse treatment facilities in the 9/13/95 NFR by running four passes of the UNDUPLICATE function in the AUTOMATCH software (Version 2.9; AUTOMATCH Technologies, Inc., Silver Spring, MD). A total of 1,520 records (about 10 percent) were identified as possible duplicates of other records on the file. This was based on a comparison of the name, address, and telephone number of the substance abuse treatment facilities. Since the NFR files are known to contain individual records for multiple treatment programs conducted at one facility location, a more extensive review of a sample of these records was carried out. This review revealed that a significant number of the records show the same name and the same telephone number, but different program identifiers. This suggests that a large proportion of the 1,520 records indeed represent multiple treatment programs within one facility, rather than true duplicates of the same facility. Therefore, these 1,520 records were retained in the ADSS sampling frame, but some may be excluded during the screening process.

It is recognized that the final ADSS sampling frame contained some ineligible records arising from (a) sampling frame errors; (b) some of the 1,520 records that turn out to be duplicates; (c) some of the 806 records with an unknown value for the NFR variable "Treatment/Prevention", which turn out to be "only Prevention" or otherwise ineligible facilities; or (d) some facilities that turn out not to have any active treatment programs (out-of-scope). The sample selection procedure was designed to account for these situations (see Section 3.8 for more details).

2.3.2 Resolution of Inconsistencies on the ADSS Frame

The resolution of inconsistencies on the ADSS frame was part of the clean-up and preparation of the sampling frame that was done prior to sample selection. Several inconsistencies were discovered when the ADSS frame was subjected to a battery of diagnostic tests. These included key-entry errors, cases with numerical codes for missing values, cases with the value of the total number of clients less than that of its component parts, and cases with percentages of various categories of clients greater than 100 percent. These inconsistencies were identified, investigated, and resolved. Most of the cases with percentage of various categories of clients greater than 100 turned out to be data-entry errors, or numerical codes for missing values that were greater than 100. These were corrected by the data processing staff. There was one case for which the number of methadone clients was far greater than the total number of clients. The sum of these two quantities was assigned to this case as the total number of clients for sampling purposes. The inconsistencies were resolved iteratively, that is, a new round of diagnostics tests was implemented after resolving each set of inconsistencies discovered in the previous round. This process continued until no more inconsistencies were found. Instead of overwriting fields in the frame file, new variables were created to reflect changes in the original variables.

2.4 Facilities Selected for the ADSS Pilot

A total of 44 out of the 46 facilities which participated in the ADSS Phase I Pilot Study exist in the ADSS sampling frame. These facilities were kept in the frame in order to ensure complete coverage. None of the pilot study facilities were selected into the ADSS sample, but 38 of them were selected into the NESAT sample.

2.5 Distribution of Facilities by State

The total number of records in the entire ADSS sampling frame is 18,368 (13,787 records from the NFR file plus 4,581 records from the enhancement file). Table 2-1 presents counts of substance abuse treatment facilities by state for the entire ADSS sampling frame and each of its components (NFR and Enhancement). Corresponding figures for the subset of the frame restricted to facilities with known and unknown number of clients are also presented. It can be seen that the percent of facilities with known number of clients varies considerably across the 50 states and the District of Columbia. Furthermore, the distribution of facilities coming from NFR and the enhancement file varies across the states.

2.6 Facility Orientation and Ownership

The variable ORIENT denotes the type of care provided by a facility, that is, whether a facility is “Drug Only”, “Alcohol Only”, or “Both Drug and Alcohol”; and the variable OWNER denotes the type of ownership of a facility. These variables are available on SAMHSA’s 1993 NDATUS survey file, and were used in the sample selection process (see Sections 3.4 and 3.6). The distributions of facilities and clients for type-of-care by sampling stratum for the entire ADSS frame and Westat’s 62 PSUs are presented in Tables 2-2a and 2-2b respectively. The corresponding distribution for type-of-ownership are given in Tables 2-3a and 2-3b.

Table 2-2a shows that of the 9,022 facilities with known type-of-care information, an overwhelming majority (7,463 or about 83 percent) of facilities have clients that are both drug and alcohol patients. Similarly Table 2-3a shows that of the 10,827 facilities by the ADSS frame with known type-of-ownership information the vast majority (6,923, or about 64 percent) are owned by private, non-profit organizations.

Table 2-2a. Distribution of type-of-care by sampling stratum in ADSS frame

	Missing				Drug				Alcohol				Drug and Alcohol				Total			
	Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients	
Stratum	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Hospital Inpatient	635	54.4	5,069	41.4	27	2.3	295	2.4	79	6.8	1,701	13.9	427	36.6	5,190	42.4	1,168	100.0	12,255	100.0
Other Residential	423	18.2	14,794	20.2	136	5.8	7,252	9.9	195	8.4	4,517	6.2	1,575	67.6	46,717	63.8	2,329	100.0	73,280	100.0
Outpatient - PM[1]	99	19.4	15,563	14.0	304	59.5	71,904	64.8	0	0.0	0	0.0	108	21.1	23,580	21.2	511	100.0	111,047	100.0
Outpatient - AEA[2]	794	38.5	64,670	31.4	0	0.0	0	0.0	366	17.7	69,231	33.6	903	43.8	72,354	35.1	2,063	100.0	206,255	100.0
Outpatient - AO[3]	2,489	40.0	204,726	38.4	378	6.1	26,538	5.0	56	0.9	9,290	1.7	3,301	53.0	292,874	54.9	6,224	100.0	533,428	100.0
Combined	1,408	54.7	126,143	48.8	9	0.3	1,220	0.5	9	0.3	2,227	0.9	1,149	44.6	128,929	49.9	2,575	100.0	258,519	100.0
Unknown	3,498	100.0	211,857	100.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	3,498	100.0	211,857	100.0
Total	9,346		642,823		854		107,209		705		86,966		7,463		569,644		18,368		1,406,642	

[1] PM = Predominantly Methadone.

[2] AEA = Almost Exclusively Alcohol.

[3] AO = All Other.

Table 2-2b. Distribution of type-of-care by sampling stratum in Westat's 62 PSUs (relevant information for Phase II)

	Missing				Drug				Alcohol				Drug and Alcohol				Total			
	Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients	
Stratum	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Hospital Inpatient	218	54.4	2,299	43.7	13	3.2	212	4.0	36	9.0	773	14.7	134	33.4	1,976	37.6	401	100.0	5,260	100.0
Other Residential	174	19.2	6,730	19.2	94	10.4	5,066	14.5	81	8.9	2,356	6.7	559	61.6	20,881	59.6	908	100.0	35,033	100.0
Outpatient - PM[1]	52	16.4	10,959	13.9	206	65.0	54,201	68.7	0	0.0	0	0.0	59	18.6	13,765	17.4	317	100.0	78,925	100.0
Outpatient - AEA[2]	283	40.4	19,319	23.1	0	0.0	0	0.0	188	26.8	42,780	51.2	230	32.8	21,505	25.7	701	100.0	83,604	100.0
Outpatient - AO[3]	1,009	40.0	87,062	38.5	249	9.9	17,435	7.7	33	1.3	6,490	2.9	1,229	48.8	114,978	50.9	2,520	100.0	225,965	100.0
Combined	430	54.8	40,901	49.2	6	0.8	811	1.0	4	0.5	494	0.6	345	43.9	40,904	49.2	785	100.0	83,110	100.0
Unknown	1,219	100.0	80,910	100.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1,219	100.0	80,910	100.0
Total	3,385		248,180		568		77,725		342		52,893		2,556		214,009		6,851		592,807	

[1] PM = Predominantly Methadone.

[2] AEA = Almost Exclusively Alcohol.

[3] AO = All Other.

Table 2-3a. Distribution of type-of-ownership by sampling strata in ADSS frame

	Missing				Private, profit				Private, non-profit				State/local government				Federal government				Tribal				Total			
	Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients	
Stratum	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Hospital Inpatient	155	13.3	3,731	30.4	167	14.3	1,931	15.8	642	55.0	3,614	29.5	170	14.6	2,146	17.5	30	2.6	833	6.8	4	0.3	0	0.0	1,168	100.0	12,255	100.0
Other Residential	228	9.8	7,404	10.1	154	6.6	3,984	5.4	1,685	72.3	47,395	64.7	254	10.9	14,307	19.5	5	0.2	153	0.2	3	0.1	37	0.1	2,329	100.0	73,280	100.0
Outpatient - PM[1]	47	9.2	9,138	8.2	143	28.0	31,488	28.4	253	49.5	57,304	51.6	66	12.9	12,889	11.6	1	0.2	213	0.2	1	0.2	15	0.0	511	100.0	111,047	100.0
Outpatient - AEA[2]	597	28.9	41,569	20.2	483	23.4	45,365	22.0	752	36.5	87,684	42.5	219	10.6	30,922	15.0	9	0.4	698	0.3	3	0.1	17	0.0	2,063	100.0	206,255	100.0
Outpatient - AO[3]	1,900	30.5	137,992	25.9	990	15.9	68,560	12.9	2,677	43.0	235,134	44.1	625	10.0	88,797	16.6	26	0.4	2,649	0.5	6	0.1	296	0.1	6,224	100.0	533,428	100.0
Combined	1,116	43.3	101,431	39.2	267	10.4	11,487	4.4	914	35.5	84,427	32.7	166	6.4	37,444	14.5	110	4.3	23,641	9.1	2	0.1	89	0.0	2,575	100.0	258,519	100.0
Unknown	3,498	100.0	211,857	100.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	3,498	100.0	211,857	100.0
Total	7,541		513,123		2,204		162,815		6,923		515,558		1,500		186,505		181		28,187		19		454		18,368		1,406,642	

[1] PM = Predominantly Methadone.

[2] AEA = Almost Exclusively Alcohol.

[3] AO = All Other.

Table 2-3b. Distribution of type-of-ownership by sampling strata in Westat's 62 PSUs (relevant information for Phase II)

	Missing				Private, profit				Private, non-profit				State/local government				Federal government				Total			
	Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients	
Stratum	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Hospital Inpatient	30	7.5	1,537	29.2	74	18.5	879	16.7	238	59.4	1,828	34.8	51	12.7	783	14.9	8	2.0	233	4.4	401	100.0	5,260	100.0
Other Residential	80	8.8	2,995	8.5	62	6.8	2,035	5.8	672	74.0	22,532	64.3	91	10.0	7,338	20.9	3	0.3	133	0.4	908	100.0	35,033	100.0
Outpatient - PM[1]	22	6.9	5,635	7.1	88	27.8	22,350	28.3	160	50.5	40,519	51.3	46	14.5	10,208	12.9	1	0.3	213	0.3	317	100.0	78,925	100.0
Outpatient - AEA[2]	202	28.8	8,943	10.7	189	27.0	25,339	30.3	250	35.7	39,411	47.1	55	7.8	9,502	11.4	5	0.7	409	0.5	701	100.0	83,604	100.0
Outpatient - AO[3]	717	28.5	52,385	23.2	454	18.0	35,697	15.8	1,137	45.1	111,259	49.2	205	8.1	25,876	11.5	7	0.3	748	0.3	2,520	100.0	225,965	100.0
Combined	300	38.2	28,531	34.3	96	12.2	3,808	4.6	310	39.5	27,357	32.9	39	5.0	12,305	14.8	40	5.1	11,109	13.4	785	100.0	83,110	100.0
Unknown	1,219	100.0	80,910	100.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1,219	100.0	80,910	100.0
Total	2,570		180,936		963		90,108		2,767		242,906		487		66,012		64		12,845		6,851		592,807	

[1] PM = Predominantly Methadone.

[2] AEA = Almost Exclusively Alcohol.

[3] AO = All Other.

3. SELECTION OF THE SAMPLE OF FACILITIES FOR ADSS PHASE I

3.1 Stratification of Facilities on the ADSS Frame

The ADSS sample design involved stratification of facilities into a number of categories. Each substance abuse treatment facility on the ADSS sampling frame was assigned to one of seven strata on the basis of its modality (the type of treatment), and other characteristics of its client population. Stratum 1 included facilities with hospital inpatient clients for both detoxification and rehabilitation. Stratum 2 included other types of active residential facilities. Stratum 3 included all outpatient facilities for which the percent of methadone clients was greater than or equal to 60 percent. Facilities assigned to stratum 4 were outpatient facilities for which the percent of alcohol-only clients was greater than or equal to 70 percent, and at the same time, the percent of methadone clients was less than 60 percent. Stratum 5 consisted of all other outpatient facilities that did not fall into stratum 3 or stratum 4. Stratum 6 included all facilities that had any other combinations of types of care defined above, but not included in the previous strata. Finally, stratum 7 included all the facilities for which no information on treatment modality and number of clients was available.

Table 1-1 shows the distribution of various facility and client statistics by stratum for the entire frame. A total of 201 facilities with unknown number of clients, which were originally in stratum 7, were restratified into strata 4 through 6 using type-of-care information, available on the sampling frame. Of these facilities, 48 were allocated to stratum 4, 122 were allocated to stratum 5, and the remaining 31 were allocated to stratum 6. This reduced the number of facilities in stratum 7 from 3,699 (see Table 2-1) to 3,498.

A comparison of these figures with those obtained for the 1993 NDATUS (see Table B1 of the Phase I OMB submission) shows significant increases in both the number of facilities and the number of clients in all strata, particularly the outpatient strata. These sharp departures from the 1993 figures are due mainly to the addition of 4,581 facilities to the frame as a result of the enhancement process.

3.2 Target Sample Sizes for ADSS Phase I

Recommendations were developed for the ADSS sample size and sample design after selecting and analyzing a number of the tables in the DSRS analytic reports. The recommendations were based on the assumption that the ADSS analytic categories would be similar to those used for DSRS. The analysis suggested that a minimum of 300 facilities were required by type of facility in order to ensure reasonably precise and stable estimates of the variables typically reported. This target minimum of 300 facilities per type of facility was interpreted as a target minimum of 300 facilities per sampling stratum since the definitions used for the DSRS type of facility and the seven ADSS sampling strata are largely interchangeable, with the necessary exception of the “Combined” and “Unknown” strata.

Under fairly mild assumptions, proportional allocation is optimal for Phase I estimates related to size. This allocation leads to less than 300 sample facilities in four of the six strata, however, and to considerably less than 300 sampled facilities in two of the strata. Determination of the actual allocation per stratum required balancing this target minimum per stratum with any associated relative increases in the design effect due to deviation from proportional allocation.

The last column of Table 1-1 gives the targeted sample sizes by stratum. As can be seen from the table, the number of clients per stratum varies greatly across strata, as does the average number of clients per facility. Under proportional allocation, the “Outpatient: Alcohol Only” and “Outpatient: All Other” strata would receive about half of the sample, while the other strata would each receive considerably less. To guarantee a minimum sample size of 300 per stratum, the “Hospital Inpatient”, “Other Residential”, “Outpatient: Predominantly Methadone” and “Combined” strata must be oversampled relative to proportional allocation. The design effect due to this oversampling is estimated at 1.50 for Phase I estimates related to size of facilities for the combined sample in the six strata.

Two of the seven NFR defined strata did not reach the minimum target of 300. This is due mostly to stratum migration. See Section 3.13 for more information about stratum migration.

3.3 Distribution of Facilities in the ADSS Frame by Size

As noted earlier, the facility measure of size is a function of the total number of clients in treatment at the facilities on October 1, 1993, for facilities on the 1995 NFR universe. Table 3-1 presents the distribution of facilities, by categories of the total number of clients in treatment, for the entire ADSS frame and each of its component parts. The corresponding figures for Westat's 62-PSU sample are also given.

Table 3-1. Distribution of facilities by categories of number of clients in treatment

	NFR				Enhancement file				Total			
	62 PSUs		Total		62 PSUs		Total		62 PSUs		Total	
Clients capacity	N	%	N	%	N	%	N	%	N	%	N	%
Missing	1,081	20.2	2,960	21.5	165	11.1	739	16.1	1,246	18.2	3,699	20.1
0	194	3.6	559	4.1	0	0.0	84	1.8	194	2.8	643	3.5
1-5	176	3.3	602	4.4	132	8.9	382	8.3	308	4.5	984	5.4
6-10	336	6.3	1,029	7.5	197	13.2	542	11.8	533	7.8	1,571	8.6
11-20	661	12.3	1,829	13.3	298	20.0	780	17.0	959	14.0	2,609	14.2
21-100	1,744	32.5	4,339	31.5	539	36.2	1,509	32.9	2,283	33.3	5,848	31.8
101-500	1,083	20.2	2,293	16.6	133	8.9	477	10.4	1,216	17.7	2,770	15.1
501+	87	1.6	176	1.3	25	1.7	68	1.5	112	1.6	244	1.3
Total	5,362	100.0	13,787	100.0	1,489	100.0	4,581	100.0	6,851	100.0	18,368	100.0

Table 3-1 shows that more than half (52 percent) of the facilities on the ADSS frame and nearly half (47 percent) of facilities located in Westat's 62 PSU sample, have less than 20 clients. Since facilities are selected in Phase I with probabilities proportional to their measure of size, such a high proportion of small facilities generally leads to extremely low probabilities of selection and, hence, extremely large sampling weights for such facilities. To avoid extreme variability in the sampling weights of facilities in the ADSS Phase I sample, a minimum measure of size was specified for facilities in the frame (see Section 3.4.2). A second reason for specifying a minimum measure of size was to guarantee that there was a sufficient pool of clients to allow the selection of the desired number of discharged records per facility in Phase II of the survey.

3.4 Measure of Size

The measure of size for records on the ADSS sampling frame is a function of the total number of clients in treatment. There were many analytical objectives, including 1) generating accurate estimates that were a function of facility characteristics, and 2) generating accurate estimates that were a function of abstract characteristics. In general, the best sample design for sampling facilities meeting objective 1) would result in equal weights of facilities (equal probability design), while the best design for objective 2) would result in unequal weights for facilities (pps design). In each of Phase I and II, both objectives were important. For ADSS, a compromise measure of size was assigned to each facility based on the distribution of research items in the Phase I analysis plan relating to facility-level and abstract-level characteristics. For multi-stage sample designs, it is common to consider the .5th power, which is called probability proportionate to the square root of size, for balancing two competing survey objectives. The resulting Phase I measure of size was a refinement of the probability proportionate to the square root of size. It became the number of clients in treatment on October 1, 1993, *x*, raised to the 0.7th power. For Phase II, the measure of size assigned to each facility was the product of the Phase I sampling interval and $x^{0.3}$. Therefore, the overall probabilities of selection for facilities for Phase II was basically:

$$p_0 = \left(\frac{n_I * x^{0.7}}{\sum x^{0.7}} \right) * \left(\frac{\frac{\sum x^{0.7}}{n_I} * n_{II} * x^{0.3}}{\sum \frac{\sum x^{0.7}}{n_I} * x^{0.3}} \right) = \frac{n_{II} * x}{\sum \frac{\sum x^{0.7}}{n_I} * x^{0.3}} .$$

The Phase II probabilities were very efficient for estimates relating to client characteristics, and eventually for the estimates resulting from the Phase II abstract analysis and Phase III client analysis. However, this was done at the expense of the estimates relating to facility characteristics in Phase II. In general, the best design for analyzing client-level data in Phase II and III would result in having equal probabilities of clients.

To assign a value for total number of clients in treatment in a given day to each of the 14,146 facilities on the NFR file, their records were linked to both the 1991 and 1993 NDATUS files, which contain information on the number of clients in treatment for each facility on the files. The 1991 data were used only when the associated 1993 record had missing data. A value for number of clients was obtained for the 10,827 NFR records which matched NDATUS records. Of the 4,581 records on the enhancement file, 3,842 records had a value for total number of clients based on screener information. Thus, out of the 18,368 records on the ADSS Sampling Frame, only 14,669 records (80 percent) had a value for the total number of clients in treatment. Since a measure of size was required for each facility prior to the selection of the sample, a measure of size was assigned to each of the 20 percent of the facilities on the sampling frame with unknown number of clients (see Section 3.4.1).

The total number of clients in treatment at a facility was used to construct the measure of size for selection of facilities in the ADSS sample. A measure of size was assigned to the 3,699 facilities for which the number of clients was unknown. The assigned measure of size was derived from all relevant information available in the ADSS sampling frame. The NFR includes information about the availability of methadone and drug and alcohol services within each facility. This information, together with the NDATUS information on the distribution of number of clients within various types of facilities was used to assign a measure of size to facilities with unknown number of clients. The procedure used to accomplish this is described in Section 3.4.1.

3.4.1 Assignment of Measure of Size to Facilities With Unknown Number of Clients

As mentioned above, a measure of size was assigned to each facility with unknown number of clients on the ADSS frame. This was done by utilizing NFR information available on the frame. Information on type of care, methadone status, client capacity, and type of ownership were initially chosen as potential variables.

3.4.1.1 Client Capacity as a Proxy for Number of Clients

Client capacity was considered as a possible predictor of the total number of clients at a facility with unknown number of clients. However, further analysis of the frame revealed that all facilities with unknown number of clients also have unknown client capacity. Table 3-2 presents a cross tabulation of various categories of number of clients in treatment, versus the same categories of client capacity for all facilities on the ADSS sampling frame with known values for both variables. This table is given here for documentation purposes only. It shows a high rate of discordance (on about 30 percent of the cases) between client capacity and number of clients. Thus, client capacity would not have been a good predictor of total number of clients in treatment, even if it was available for facilities with unknown number of clients.

Table 3-2. Cross tabulation of client capacity versus number of clients in treatment

	Number of clients in treatment							
Clients capacity	0	1-5	6-10	11-20	21-100	101-500	501+	Total
0	1	0	0	0	0	0	0	1
1-5	27	120	4	1	0	0	0	152
6-10	36	175	355	28	2	0	0	596
11-20	43	143	401	922	32	0	0	1,541
21-100	42	128	233	798	3,571	47	0	4,819
101-500	4	6	9	31	650	2,095	7	2,802
501+	0	1	0	0	21	133	164	319
Total	153	573	1,002	1,780	4,276	2,275	171	10,230

3.4.1.2 Type-of-Care Information

Prior to finalizing the procedure for assigning measures of size to facilities with unknown number of clients, the reliability of information from the NFR file was evaluated by comparing the distribution of facilities by orientation and methadone status using NFR information with the corresponding distribution using NDATUS information. The results are presented in Tables 3-3a and 3-3b.

Table 3-3a. Number of drug and/or alcohol facilities in the subset of ADSS frame with NDATUS information on type-of-care

	NDATUS			
NFR	Drug	Alcohol	Drug and Alcohol	Total
Drug	851	1	0	852
Alcohol	0	703	0	703
Drug and Alcohol	3	1	7,463	7,467
Total	854	705	7,463	9,022

Table 3-3b. Number of methadone facilities in ADSS frame

	NDATUS		
NFR	Drug	Alcohol	Total
Methadone	10,188	548	10,736
Non-methadone	3	88	91
Total	10,191	636	10,827

Note that the figures in the above tables are based on facilities with the desired information from both sources of data. The results in Tables 3-3a and 3-3b show that NFR information is reliable in classifying facilities as drug and/or alcohol, but not as reliable in classifying facilities as methadone/non-methadone. It should also be noted that all facilities with unknown number of clients on the ADSS frame also have no information on type of ownership. These findings were taken into consideration in the assignment of differential selection probabilities to different types of facilities such that the final sampling weights of facilities of the same modality are as close as possible, regardless of whether they are from stratum 7 or from the first six strata.

3.4.1.3 Derivation of a Measure of Size

In order to give every facility on the ADSS sampling frame a probability of selection and, at the same time, guard against widely varying sampling weights in the final ADSS sample, a measure of size was assigned to all facilities with unknown number of clients. This was done in two different ways, depending on the stratum of such facilities. For the 3,498 facilities in stratum 7, where the total number of clients in treatment is unknown for all facilities, the facilities were partitioned into a number of substrata based on available NFR information.

Information relating to whether a facility administers methadone treatment in NFR was found not to be useful because it is not reliable (see Table 3-5b), and it is available only for about 0.1 percent of the facilities in stratum 7. As indicated above, the most useful NFR information is the type-of-care variable, which classifies facilities according to whether they are "Drug Only", "Alcohol Only", and "Drug and Alcohol". This NFR variable was found to have an acceptable quality based on an evaluation of it against similar information provided for the NDATUS respondents (see Table 3-3a). The above categories of the variable were then used to form the substrata for stratum 7. Facilities with no information from NFR were classified under the substrata labeled "Unknown - No Information". Table 3-3c gives the distribution of facilities over the four substrata. Note that of the 3,498 facilities in stratum 7, only 538 facilities have no NFR information.

Table 3-3c. Distribution of facilities over the four substrata of stratum 7

Substratum	Description	Number of facilities
7a	Unknown - Drug	94
7b	Unknown - Alcohol	137
7c	Unknown - Drug and Alcohol	2,729
7d	Unknown - No information	538
Total		3,498

A value for the number of clients was assigned to each of the facilities equal to the average number of clients for the type indicated, taking into account such factors as census region, metro status, and facility orientation. For the 538 facilities in the "Unknown-No NFR Information" category, an overall average number of clients was assigned. For the 201 facilities with unknown number of clients in strata 4, 5, and 6, the stratum-based average number of clients of each type was assigned.

3.4.2 The Minimum Number of Clients Associated With a Facility

A substantial number of facilities on the ADSS frame have less than five clients in treatment (see Table 3-2), even after assigning number of clients to the 3,699 facilities with unknown number of clients. A minimum measure of size was established for each facility on the ADSS frame for sampling purposes. This was done to avoid extreme variability in the sampling weights of facilities in the ADSS Phase I sample. For sampling purposes, the minimum number of clients was set to 3 for all facilities with no clients in stratum 1 (hospital inpatient facilities) and to 5 for all other facilities in the sampling frame. The minimum number of clients was defined differently for stratum 1 because nearly half of the facilities in this stratum had no clients. Assigning a minimum number of clients value of 5 to these facilities would probably have resulted in an inordinate number of sampled facilities from this subgroup of stratum 1. Thus, the final number of clients in treatment used for sampling was set equal to the total number of clients in treatment at the facility if it was at least 5. If the total number of clients in treatment at a facility was less than 5, then the final number of clients depended on the stratum. For stratum 1, it was set to 3 if it facility had no clients, and 5 if it had at least 1 client. For all other strata, it was set to 5.

As mentioned earlier, the ADSS sample selection was designed to minimize overlap with the NESAT survey. A brief overview of the NESAT sample design is given in the following section. Section 3.8 includes the details of the sample design used to minimize overlap between the ADSS and NESAT surveys.

3.5 Brief Overview of the NESAT Sample Design

The NESAT survey has a multi-stage design. The first stage sample is Westat's national area sample of 62 PSUs, restricted to those PSUs located in metropolitan areas in the United States, 50 in all. The decision to restrict the NESAT survey to metropolitan areas was based on the assumption that most substance abuse treatment facilities are located in metropolitan areas and concentrating on these areas would allow for the most productive utilization of the survey resources. The 50 PSUs are a subset of the Westat Master sample, consisting of the 24 metro certainty PSUs and the 26 metro non-certainty PSUs in the half-sample of metro non-certainty PSUs not in Westat's 62 PSU sample. The second stage involves the selection of a screening sample of substance abuse treatment facilities within sampled PSUs. This sample was selected for the purpose of identifying the NESAT-eligible Service Delivery Units (SDUs) associated with each sampled facility. The third stage involved the selection of a national sample of about 200 substance abuse treatment SDUs, around 40 within each of five separate modalities. The five modalities are: inpatient, methadone, Therapeutic Communities (TCs), intensive outpatient treatment, and non-intensive outpatient treatment. The fourth stage involved the selection of adult clients within the sampled SDUs.

As mentioned earlier, the ADSS survey also has a multi-stage stratified design. To minimize overlap between ADSS and NESAT and, at the same time, increase the sample yield in the non-certainty areas of the PSU sample, a first stage sample of facilities was selected from the entire ADSS sampling frame, with the exception of the 26 metro non-certainty PSUs in the NESAT sample, and with the sampling rates in the non-certainty areas being twice as large as those in the remainder of the sampling frame. At the second stage, a sample of client discharge records will be selected from facilities responding to Phase I of the survey, which are located in Westat's 62-PSU sample.

As indicated above, the choice of PSUs was made to minimize the overlap between ADSS and NESAT. This choice guarantees that overlap can occur only in the 24 metro certainty PSUs. Section 3.6 describes the methodology used to minimize overlap within the 24 metro certainty PSUs.

3.6 Phase I Sample Selection

The selection of the sample of facilities for Phase I of ADSS was carried out in conjunction with sample selection for the NESAT survey. Sample selection was done in such a way as to minimize the overlap between the two surveys, while still employing sample designs that are consistent with the objectives of the two surveys. This section gives details of the sample selection procedures followed in certainty and non-certainty areas of the sampling frame.

By design, all of the potential overlap between the ADSS and NESAT surveys was expected to be in the 24 metro certainty PSUs (see Appendix A). To minimize overlap between the two surveys, different sample selection procedures were undertaken in the certainty and non-certainty areas of the sampling frame. A description of the sample selection procedures is given in Section 3.6.1 for the certainty PSUs and in Section 3.6.2 for the non-certainty PSUs. The evaluation of the potential overlap between ADSS and NESAT is presented in Section 3.6.1.1.

Table 3-4a presents the distribution of facilities and clients across the ADSS sampling strata for ADSS and NESAT PSUs, prior to the assignment of number of clients to the facilities with unknown number of clients (see Section 3.4.1). Table 3-4b presents the distribution of facilities and their associated measure of size (after assignment of number of clients to facilities with unknown number of clients) by sampling stratum for the entire frame, partitioned into various subsets (metro certainty PSUs, metro non-certainty PSUs divided into ADSS PSUs, that is, Westat's 62 PSUs; and NESAT PSUs, the remaining non-MSAs, and the remainder, consisting of PSUs in the nation, outside the Westat Master

Table 3-4a. Distribution of facilities and clients by sampling stratum before imputation of number of clients

	Certainty MSAs				Non-Certainty MSA PSUs												Non-MSA PSUs				Remainder				Total			
					ADSS PSUs				NESAT PSUs				Total PSUs															
	Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients		Facilities		Clients	
Stratum	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Hospital Inpatient	327	28.0	3,814	31.1	72	6.2	1,431	11.7	97	8.3	939	7.7	169	14.5	2,370	19.3	8	0.7	69	0.6	664	56.8	6,002	49.0	1,168	100.0	12,255	100.0
Other Residential	716	30.7	28,628	39.1	189	8.1	6,318	8.6	221	9.5	5,702	7.8	410	17.6	12,020	16.4	13	0.6	239	0.3	1,190	51.1	32,393	44.2	2,329	100.0	73,280	100.0
Outpatient - PM[1]	282	55.2	70,723	63.7	35	6.8	8,202	7.4	48	9.4	11,980	10.8	83	16.2	20,182	18.2	0	0.0	0	0.0	146	28.6	20,142	18.1	511	100.0	111,047	100.0
Outpatient - AEA[2]	569	27.6	73,097	36.2	127	6.2	9,470	4.7	147	7.1	13,786	6.8	274	13.3	23,256	11.5	10	0.5	614	0.3	1,210	58.7	104,863	52.0	2,063	100.0	201,830	100.0
Outpatient - AO[3]	1,906	30.6	171,266	32.7	594	9.5	52,223	10.0	502	8.1	42,917	8.2	1,096	17.6	95,140	18.2	44	0.7	3,246	0.6	3,178	51.1	253,695	48.5	6,224	100.0	523,347	100.0
Combined	598	23.2	59,180	23.2	183	7.1	23,200	9.1	210	8.2	21,983	8.6	393	15.3	45,183	17.7	14	0.5	776	0.3	1,570	61.0	150,411	58.9	2,575	100.0	255,550	100.0
Unknown	936	26.8	0	0.0	278	7.9	0	0.0	304	8.7	0	0.0	582	16.6	0	0.0	23	0.7	0	0.0	1,957	55.9	0	0.0	3,498	100.0	0	0.0
Total	5,334		406,708		1,478		100,844		1,529		97,307		3,007		198,151		112		4,944		9,915		567,506		18,368		1,177,309	

[1] PM = Predominantly Methadone.

[2] AEA = Almost Exclusively Alcohol.

[3] AO = All Other.

Table 3-4b. Distribution of facilities and final measure of size (MOS_F) by sampling stratum

	Certainty MSAs				Non-Certainty MSA PSUs												Non-MSA PSUs				Remainder				Total			
					ADSS PSUs				NESAT PSUs				Total PSUs															
	Facilities		MOS_F		Facilities		MOS_F		Facilities		MOS_F		Facilities		MOS_F		Facilities		MOS_F		Facilities		MOS_F		Facilities		MOS_F	
Stratum	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Hospital Inpatient	327	28.0	1,712	29.9	72	6.2	454	7.9	97	8.3	466	8.1	169	14.5	920	16.1	8	0.7	38	0.7	664	56.8	3,050	53.3	1,168	100.0	5,720	100.0
Other Residential	716	30.7	8,164	35.6	189	8.1	1,946	8.5	221	9.5	1,984	8.6	410	17.6	3,930	17.1	13	0.6	97	0.4	1,190	51.1	10,747	46.9	2,329	100.0	22,938	100.0
Outpatient - PM[1]	282	55.2	12,814	61.7	35	6.8	1,523	7.3	48	9.4	2,144	10.3	83	16.2	3,667	17.7	0	0.0	0	0.0	146	28.6	4,287	20.6	511	100.0	20,768	100.0
Outpatient - AEA[2]	569	27.6	14,565	33.6	127	6.2	2,187	5.0	147	7.1	3,094	7.1	274	13.3	5,281	12.2	10	0.5	175	0.4	1,210	58.7	23,372	53.9	2,063	100.0	43,393	100.0
Outpatient - AO[3]	1,906	30.6	37,546	31.5	594	9.5	11,765	9.9	502	8.1	10,133	8.5	1,096	17.6	21,898	18.4	44	0.7	798	0.7	3,178	51.1	58,833	49.4	6,224	100.0	119,075	100.0
Combined	598	23.2	12,575	23.1	183	7.1	4,607	8.5	210	8.2	4,742	8.7	393	15.3	9,349	17.2	14	0.5	213	0.4	1,570	61.0	32,272	59.3	2,575	100.0	54,409	100.0
Unknown	936	26.8	17,830	29.0	278	7.9	4,909	8.0	304	8.7	5,541	9.0	582	16.6	10,450	17.0	23	0.7	349	0.6	1,957	55.9	32,939	53.5	3,498	100.0	61,568	100.0
Total	5,334		105,206		1,478		27,391		1,529		28,104		3,007		55,495		112		1,670		9,915		165,500		18,368		327,871	

[1] PM = Predominantly Methadone

[2] AEA = Almost Exclusively Alcohol

[3] AO = All Other

Sample). Within each stratum, a systematic sample with a fixed target sample size (about twice as large as the target numbers in Table 1-1) was drawn with probability proportional to a measure of size equal to the 0.7-th power of the final number of clients in treatment. The number of facilities allocated to each subset of the frame in each stratum was proportional to the distribution of the total measure of size over the subsets in each stratum, as given in Table 3-4b.

The resulting allocations of the screening samples of facilities to the various strata for ADSS and NESAT are presented in Table 3-5.

Table 3-5. Expected sample allocation across the ADSS sampling strata by PSU type for the ADSS and NESAT screening samples

	ADSS			NESAT		
Stratum	Certainty PSUs	Remainder of the nation	Total	Certainly PSUs	Metro non-certainty PSUs	Total
Hospital Inpatient	179	425	604	24	28	52
Other Residential	214	388	602	164	191	355
Outpatient - PM[1]	282	229	511	125	126	251
Outpatient - AEA[2]	202	400	602	103	132	235
Outpatient - AO[3]	316	694	1,010	279	336	615
Combined	139	463	602	89	107	196
Unknown	230	586	816	100	100	200
Unknown - Drug	19	5	24			
Unknown - Alcohol	19	21	40			
Unknown - Drug and Alcohol	167	461	628			
Unknown - No Information	25	99	124			
Total	1,562	3,185	4,747	884	1,020	1,904

[1] PM = Predominantly Methadone.

[2] AEA = Almost Exclusively Alcohol.

[3] AO = All Other.

3.6.1 Sample Selection Within Certainty PSUs

As mentioned above, overlap between ADSS and NESAT was limited to the metro certainty PSUs by design. The sample selection procedure was thus designed to minimize the overlap in the certainty PSUs. The probability of selecting a facility for inclusion in the ADSS screener sample is denoted by the variable *PROB_ADS*. For facilities in certainty PSUs, the probability of selection of the *j*-th facility in the *i*-th stratum into the screening sample is

$$PROB\_ADS = \frac{A_i * ADSSMOS_{ij}}{\sum_j ADSSMOS_{ij}};$$

where A_i is the target sample size for stratum *i*, and $ADSSMOS_{ij}$ is the measure of size for the *j*-th facility in the *i*-th stratum (the total number of clients in treatment, raised to the power 0.7). The corresponding probability of selection for the NESAT sample is denoted by *PROB_NES*. A facility is designated as a certainty facility if its probability of selection is greater than or equal to 1.

Certainty selections were identified on the basis of these initial probabilities of selection. They were then removed from the sampling process and the total number of facilities to be sampled from the associated stratum was reduced by 1 for each such certainty selection. Furthermore, the total measure of size for the stratum was reduced by the amount accounted for by the certainty selections, and the probabilities of selection for the remaining facilities were recomputed. This procedure was implemented iteratively until no certainty selections remain or all facilities within a stratum are selected with certainty.

In order to minimize the degree of overlap, the Permanent Random Number approach was used to select facilities in the certainty PSUs for both ADSS and NESAT. The Permanent Random Number approach provides a simple and straightforward method of minimizing overlap, and it is applicable even when two surveys use different measures of size, as was the case for ADSS and NESAT. The approach was implemented as follows: First, a variable *PRN* was created, assigning a random number from the Uniform (0, 1) distribution, to each facility in each stratum. This is the permanent number associated with each facility. The values of *PRN* were then compared to the probabilities of selection of each facility into the ADSS and NESAT samples. The selection of a facility for inclusion into the ADSS or NESAT sample depended on the relationship between *PRN* and the probabilities of selection. The facilities were selected as follows:

- If 0≤*PRN*≤*PROB_ADS*, then the corresponding facility was selected for the ADSS sample; and
- If (1 - *PROB_NES*)≤*PRN*≤1, then the corresponding facility was selected for the NESAT sample.

The resulting sample has the following characteristics:

- Only the facilities in the 24 metro certainty PSUs had a chance of overlapping between the initial samples of the two surveys;
- There was no overlap in the initial samples of the two surveys for those facilities in the metro certainty PSUs where the probabilities of selection for both surveys were less than 0.5; and
- Those facilities with probabilities of selection in both surveys greater than or equal to one were in the initial samples of both surveys with certainty.

For more details on the Permanent Random Number approach, refer to Ohlsoon (1995).

3.6.1.1 Overlap Between the ADSS and NESAT Surveys in Certainty PSUs

As mentioned earlier, only facilities selected from the 24 certainty PSUs were subject to possible overlap. Based on their number of facilities and the corresponding client size, the 24 certainty PSUs used for both ADSS and NESAT should have accounted for 761 facilities in the ADSS survey. The extent of overlap between the two surveys was estimated on the basis of the information about the facilities available in the sampling frame, the allocation of the NESAT and the ADSS samples across the strata, and the specific features of the design for each of the surveys.

The expected overlap between ADSS and NESAT was evaluated at the end of the screening of the ADSS sample of facilities. A total of 2,771 facilities completed the ADSS screener questionnaire and were eligible for ADSS Phase I. Similarly, the NESAT sample of SDUs is expected to consist of 200 eligible service delivery units (SDUs) or programs. The following is a summary of the results of a preliminary evaluation of the extent of true overlap between the ADSS Phase I sample of 2,436 facilities and the NESAT sample of 200 program. The estimated overlap between the ADSS screened sample of 2,771 facilities and the NESAT sample of 200 programs was computed as the maximum possible overlap between ADSS Phase I and NESAT. The true overlap between ADSS Phase I and NESAT will be known after data collection is completed for both surveys. Since it is possible for several programs to be associated with the same facility, the evaluation of overlap was done at the facility level.

A total of 21 records were found to be common to both the NESAT SDU file and the ADSS Phase I facilities file. Two of the records appeared to represent additional programs within overlapping facilities. Also, one of the overlapping SDUs did not share the same location address as the sampled facility associated with it. This SDU was therefore not considered as part of the ADSS/NESAT overlap.

Therefore, the maximum number of truly overlapping facilities between the ADSS screened sample and the NESAT sample of programs is 18. The majority (11) of the overlapping facilities are in the methadone stratum. There is no overlap in stratum 4. The distribution of the overlapping facilities in the ADSS screened sample and the NESAT sample is given in Table 3-6 below.

Table 3-6. Actual overlap between the ADSS screened sample of 2,771 facilities and the NESAT sample of 200 programs by analytic stratum (this is the maximum possible overlap between ADSS Phase I and NESAT)

Analytic stratum	Number of overlapping facilities
1	1
2	3
3	11
4	0
5	1
6	2
Total	18

3.6.2 Sample Selection Within Non-certainty Areas of the Nation

Within the non-certainty areas of the nation, the ADSS sample was selected by systematic sampling within each of the seven sampling strata separately. As mentioned earlier, the 26 NESAT metro non-certainty PSUs were avoided. To accomplish this, the measure of size variable for ADSS was set to zero for the NESAT metro non-certainty PSUs, and was doubled for the ADSS metro non-certainty PSUs. This procedure helped accomplish two desirable objectives: increasing the sample yield in ADSS for the Phase II sample selection within Westat's 62 PSU area sample, while avoiding overlap with the NESAT sample in all but the metro certainty PSU areas.

The ADSS sample was selected by using the random systematic selection procedure available in WESSAMP (a system of macros developed by Westat for sample selection). Before sample selection, the records on the sampling frame were first sorted by type of PSU (metro certainty; metro non-certainty; non-metro, non-certainty), census region, membership in Westat's 62-PSU area sample, the PSU, type of ownership, and the final number of clients at a facility.

At the conclusion of the sample selection process, a number of diagnostic checks were performed to evaluate the quality of the selected sample. In particular, the number of sampled records was examined for each PSU for each stratum separately, and for each PSU across all strata. For each stratum, the distribution of the variable denoting the total number of clients in treatment at a facility for the set of sampled records was compared to that of the corresponding variable actually used for sample selection purposes. The distribution of the actual sample of 4,691 facilities across all the strata is given in Table 3-7 below.

Table 3-7. The distribution of the actual ADSS screening sample by sampling stratum and type of PSU

	Number of records in ADSS screening sample		
Stratum	Certainty PSUs	Non-certainty PSUs	Total
Hospital Inpatient	178	425	603
Other Residential	212	388	600
Outpatient - PM[1]	282	181	463
Outpatient - AEA[2]	198	400	598
Outpatient - AO[3]	331	694	1,025
Combined	132	463	595
Unknown	221	586	807
Total	1,554	3,137	4,691

[1] PM = Predominantly Methadone.
[2] AEA = Almost Exclusively Alcohol.
[3] AO = All Other.

Since the responding facilities located in Westat's 62 PSUs form the sampling frame for Phase II of ADSS, the distribution of sampled facilities in this area sample of PSUs was also examined. The results revealed that:

- Facilities were selected from all 24 certainty PSUs;
- Of the 38 non-certainty PSUs, facilities were selected in 33 PSUs; and

- Of the remaining 5 non-certainty PSUs only two PSUs contained substance abuse treatment facilities and none of them were selected into the screening sample. The remaining three PSUs had no facilities located in them.

3.7 Partitioning of the ADSS Screening Sample Into Two Waves

For Phase I of ADSS, fixed sample sizes were required for each of the six types of modality in the first six strata. For all strata, sampling rates were specified on the basis of the distribution of facilities in the frame, and the sample sizes required for ADSS. These sampling rates were then used to select the Phase I sample of facilities. The derivation of these rates were based on a number of assumptions which were expected to hold only approximately. For example, it was assumed that the response rate is uniform across different modalities, and a specific distribution (across modality) was assumed for facilities for which the sampling information is not known (for instance, facilities in stratum 7). Also, a uniform rate of out-of-scope (e.g., private practitioners) and duplicates was assumed across all strata. The composition of the final sample within the sampling strata would have been similar to the required composition only to the extent the above assumptions proved to be true.

There was no way of knowing in advance whether the assigned sampling rate for a particular modality would produce a higher or lower number of facilities in that stratum. Consequently, a sample selection procedure was necessary that could produce samples either larger or smaller than that produced by a fixed sampling rate. One approach that is commonly used when fixed sample sizes are required is to conduct sequential sampling. The sequential sampling procedure used for screening the ADSS sample is described below.

The first step in the sequential sampling procedure was the selection of a screening sample that was twice as large as the target sample sizes for Phase I, both at the stratum level and overall. The screening sample was then partitioned into two subsamples or waves.

The initial sample of records for 4,691 facilities was partitioned into two waves within each of the sampling strata for the sequential release of the screening sample. The first wave was released for screening. The second wave of facilities was set aside and not used until screening was completed in the first wave. At the conclusion of screening the first set, frequencies of out-of-scope, duplicates, response rates, and other sources of attrition were used to update the sampling rates within the seven strata such that the expected final sample size was close to the target numbers for Phase I of ADSS. A sample of facilities was then selected from the second half-sample using the revised rates. Sequential sampling resulted in facility sample sizes that are closer to the targeted numbers by modality. However, these sample sizes were achieved at the price of variable sampling rates introduced by sequential sampling. A summary of the revised sample sizes in given in Section 3.8. This section describes how sequential sampling was used in Phase I.

The assignment of facilities to the two waves was done in such a way as to preserve the appropriate probability of selection of each facility into the ADSS sample, which is only half the size of the screener sample selected. The goal was to select facilities into the ADSS sample with probabilities proportional to their measures of size. Assuming that the screener sample is exactly twice as large as the target ADSS sample size, not only overall, but also within each strata, the variable *WAVEPROB* was defined as

$$WAVEPROB = \frac{PROB\_ADS}{2}.$$

The variable *WAVEPROB* represents the probability of selection of each facility in the ADSS Phase I sample if only half of the screening sample (the first wave) was selected for screening. Note that if $WAVEPROB \geq 1$, then the corresponding facility would have been selected with certainty in the screener sample consisting of both waves. If $0.5 \leq WAVEPROB < 1$, then the corresponding facility would have been selected with certainty for the screener sample, but not for the first wave. If $WAVEPROB < 0.5$, then the corresponding facility would not have been selected with certainty in either screener sample or the first wave. Facilities in the screener sample were then assigned to the two waves as follows:

1. The few sampled facilities in the non-MSA PSUs in Westat's 62 PSU sample were assigned to Wave 1 to preserve the sample sizes for non-MSAs in Phase II of the survey;

2. Sampled facilities with $WAVEPROB \geq 1$ were assigned to Wave 1; and

3. Sampled facilities with $WAVEPROB < 1$ were first divided into two groups: facilities with $0.5 \leq WAVEPROB < 1$ and facilities with $WAVEPROB < 0.5$. For each group, the facilities were sorted by the sample selection order (metro status, census region, PSU, type of ownership, and final number of clients). The sampled facilities within each group were then assigned to the two waves alternately, that is, with probability 0.5 for each wave.

The above procedure for assigning sampled facilities to waves ensures that facilities are assigned to the first wave with the appropriate probability of selection into the ADSS sample.

Table 3-8 gives the distribution of the various categories of sampled facilities by sampling stratum. Various listings of sample records in both waves were examined to ascertain that all sampling variables, especially the probabilities of selection, are correctly assigned to the waves.

Table 3-8. Distribution of sampled facilities by sampling stratum

Stratum	Number of facilities with *WAVEPROB*≥1	Number of facilities with 0.5≤*WAVEPROB*<1	Number of facilities with *WAVEPROB*<0.5	Total
Hospital Inpatient	7	224	372	603
Other Residential	1	49	550	600
Outpatient - PM[1]	51	412	0	463
Outpatient - AEA[2]	1	113	484	598
Outpatient - AO[3]	2	54	969	1,025
Combined	2	83	510	595
Unknown - Drug	0	0	23	23
Unknown - Alcohol	0	0	39	39
Unknown - Drug and Alcohol	0	0	621	621
Total	64	935	3,692	4,691

[1] PM = Predominantly Methadone.
[2] AEA = Almost Exclusively Alcohol.
[3] AO = All Other.

The distribution of sampled facilities across the sampling strata for each of the two waves is presented in Table 3-9. Of the 4,691 records selected for screening, 2,385 records were allocated to Wave 1 and 2,306 were allocated to Wave 2. Table 3-9 shows that the sampled records are about evenly distributed across the two waves for each stratum and overall.

Table 3-9. Distribution of sampled facilities by sampling stratum for the two screening waves

Stratum	Wave 1	Wave 2	Total
Hospital Inpatient	307	296	603
Other Residential	301	299	600
Outpatient - PM[1]	257	206	463
Outpatient - AEA[2]	301	297	598
Outpatient - AO[3]	516	509	1,025
Combined	299	296	595
Unknown	404	403	807
Total	2,385	2,306	4,691

[1] PM = Predominantly Methadone.
[2] AEA = Almost Exclusively Alcohol.
[3] AO = All Other.

3.8 The Process of Selection of Facilities for ADSS Phase I

As already mentioned, sample selection for Phase I of ADSS was carried out in two stages. First, a sample of records, approximately twice the size required for Phase I, was selected from the ADSS sampling frame. Second, the selected records were screened sequentially in two waves. All the facilities assigned to Wave 1 were screened for participation in ADSS. However, only a subsample of facilities assigned to Wave 2 were screened. Table 3-11 presents the frequencies related to the screening of facilities for ADSS. Responding facilities which were eligible for Phase I during screening were included in the Phase I sample.

The provisional ADSS screener sample consisted of 4,189 records, and its distribution is shown in Table 3-10 below.

Table 3-10. The provisional ADSS screener sample

Type of records	Number
Records screened in Wave 1	2,385
Records selected for Wave 2	1,621
Supplemental screener sample for Stratum 4	89
Facilities administered by administrative units	93
Extra record added to screener file (see Section 3.12 below)	1
Total	4,189

Of the 4,189 facilities in the screener sample, a total of 589 facilities were deselected from the sample because their corresponding strata had the sample size necessary to obtain the number of completed cases possible. The vast majority of these facilities were not screened. However, about 43 facilities slated for de-selection were screened before the de-selection was implemented. In addition, two facilities administered by de-selected administrative units were screened before their de-selection. These 45 facilities were tracked to ensure that their Phase I weights were adjusted (their weights will be set to 1). See Section 3.9 for details.

A total of 2,771 facilities responded to the screener and were eligible for Phase I of ADSS. Our goal at this point was to construct a unique set of facilities sampled for Phase I, compute their probabilities of selection, and construct their base weights.

Table 3-11. NFR strata frequencies used to select ADSS Phase I sample

	In NFR frame		Oversample (chosen using PPS)		Wave 1[*]		Wave 2[**]		Screened in Wave 2		Total screened	
NFR Strata	N	%	N	%	N	%	N	%	N	%	N	%
Hospital Inpatient	1,168	6.4	603	12.9	307	12.9	296	12.8	296	26.4	603	17.2
Residential	2,329	12.7	600	12.8	301	12.6	299	13.0	221	19.7	522	14.9
Outpatient-Methadone	511	2.8	463	9.9	257	10.8	206	8.9	206	18.4	463	13.2
Outpatient-Alcohol	2,063	11.2	598	12.7	301	12.6	297	12.9	297	26.5	598	17.1
Outpatient-Other	6,224	33.9	1,025	21.9	516	21.6	509	22.1	1	6.1	517	14.7
Combined	2,575	14.0	595	12.7	299	12.5	296	12.8	50	4.5	349	10.0
Unknown	3,498	19.0	807	17.2	404	16.9	403	17.5	50	4.5	454	12.9
Total	18,368	100.0	4,691	100.0	2,385	100.0	2,306	100.0	1,121	100.0	3,506	100.0

[*] All facilities in Wave 1 were screened for participation in ADSS.

[**] Only some facilities in Wave 2 were screened for participation in ADSS.

3.9 Deselected Units Remaining in the Sample

There were 45 facilities that were inadvertently included in the Phase I sample. These facilities were to be deselected from Wave 2 of the screener. Two of these facilities were children of an administrative unit, which was a deselected facility. Twelve of the remaining 43 facilities became Phase I respondents (1 in analytic stratum 2, nine in stratum 5, and two in stratum 6). Seven of the 12 Phase I respondents that were deselects inadvertently became eligible for selection in Phase II and three were sampled for the incentive study. Each of the deselects that were sampled in Phase I were assigned a base weight of one.

3.10 Adding an Extra Record to Phase I

During screening, a facility was identified as having split off from a sample facility. The identified facility was not on the sampling frame prior to sample selection, and hence, was not sampled or released for screening. It was therefore added to the Phase I files. A Phase I questionnaire was subsequently administered to this facility. The data for the original facility was filled in for the added facility for all variables except the disposition codes. In particular, the added facility was given the probability of selection associated with the original sample facility in the case of the added facility.

3.11 Identification of Facilities With Multiple Chances of Selection

There was strong evidence during the construction of the ADSS sampling frame suggesting that it contained many potential duplicates. Thus it was expected that some of the facilities in the ADSS screener sample would have multiple chances of selection. An attempt was made to identify these facilities by comparing the set of records in the screener sample to the ADSS sampling frame. This was done in two ways: by manual look-ups during screening; and by employing the GEOMATCH/DUPLICATES program in the AUTOMATCH software.

3.11.1 Duplicates Identified During Screening

One or more duplicates were identified on the sampling frame during the screening of each of a number of sampled records. The search for duplicates was conducted whenever the sampled facility reported changes in key characteristics (name, location address, telephone number, type of care, etc.) from those characteristics used for sampling (based on NFR information). If, based on the new information, the sampled record was found to match at least one other record on the sampling frame, the sampled record and the associated frame record were declared duplicates and included in a list for the purpose of adjusting the probabilities of selection of the sampled records. A total of 65 records were found to be duplicates of other records, the majority (62 or 95 percent) of which were on the sampling frame. The remaining three duplicates were in the sample. The duplicates identified within the sample were coded as ineligible and were excluded from screening activities.

3.11.2 Duplicates Identified by AUTOMATCH

The AUTOMATCH software was used to match the screener-sampled records to records on the sampling frame. The matching process utilized such pieces of information as facility name, program name, full location address (including city, state, and zip code), and telephone number. It was assumed throughout the matching process that each unique pair of location address and telephone number represents a single facility.

The GEOMATCH DUPLICATES program in AUTOMATCH was run in three passes at once, with each pass run on the residuals of the preceding pass. The passes were in order from the most restrictive to the least restrictive. A list of potential duplicates identified by the three passes of AUTOMATCH was compiled. For each sampled record, a set of records on the ADSS sampling frame that share the same facility name, program name, and location address was compiled.

The list of potential duplicates identified by AUTOMATCH was manually reviewed by the statisticians. Records with the same facility name, program name, location address, and telephone number were declared as duplicates. Telephone number was considered the least reliable among the variables examined to identify duplicates (true duplicates may have different telephone numbers). Records which did not meet the criteria for true duplicates were eliminated. At the end of the manual examination of the list of potential duplicates identified by AUTOMATCH was compiled.

A final file containing sampled records and their true duplicates in the ADSS sampling frame, identified by both the manual look-up and AUTOMATCH, was created. This file contained a total of 168 sampled records each associated with one or more duplicates in the ADSS sampling frame, along with all the information necessary for the adjustment of the probabilities of selection of the sampled records. The distribution of sampled records with duplicates is given in Table 3-12 below.

Table 3-12. Distribution of sampled records with duplicates

Number of duplicates	Number of sampled records
1	149
2	16
3	2
>3	1
Total	168

In addition, during data collection, 103 sampled facilities were identified as duplicates of other sampled facilities.

3.11.3 Computing the Appropriate Probabilities of Selection

The final probability of selection of each record in each ADSS stratum (for the sample of records released for screening) was determined and assigned to all records on the frame. This probability of selection reflects the original probability of selection into the screener sample, and the subsampling of the sample selected for Wave 2. Each unique sampled record was then associated with a set of duplicates on the sampling frame, if such duplicates existed. The identification of the duplicates was described in Section 3.11.2. The final probabilities of selection for the sampled records were adjusted to account for duplicate records that were identified.

3.12 Administrative Units

During the ADSS screener, some sampled facilities were identified as being administrative units for other facilities called children facilities. There were 28 administrative units identified by the screener questionnaire. Two of the 28 administrative units requested that one questionnaire be filled out for all its associated children. The child facilities of the two administrative units were coded as ineligible. The eligibility status of the two administrative units was based on the status of the administrative units itself. One of the administrative units had three child facilities, and the other had 13 child facilities. For the remaining 26 parent facilities, there were 77 child facilities. The probability of selection of the parent facility was given to its associated child facilities. Each child facility received a Phase I questionnaire. A total of 93 facilities were identified as eligible "children" of administrative units and were added to the sample. Of these, three were identified as duplicates of other facilities, and were coded as ineligible.

The final probability of selection of each administrative unit was assigned to each facility identified by an administrative unit. All other variables involved in the selection of the screener sample (except the number of clients in treatment), were assigned in the same way (the values of the variables for an administrative unit were assigned to all facilities identified by it).

For Phase II, only completed cases in the Westat 62 PSUs were considered eligible. In addition, the 26 parent facilities mentioned above were not eligible and the 16 children facilities associated with the two responding parent facilities that filled out one questionnaire were not eligible.

3.13 Stratum Migration

During the Phase I interview, a number of facilities were found to have changed treatment modalities from those to which they were assigned prior to sample selection. These assignments were based on information available on the ADSS sampling frame. Table 3-13 clearly shows a very high level of stratum migration between sample selection and the Phase I interview. As an extreme example, of the facilities sampled in stratum 4, only 30.73 percent remained in stratum 4, and 66.34 percent migrated to stratum 5. Stratum 5 grew from 389 facilities (among Phase I respondents) to 891 facilities. Most of the migration came from stratum 4. Consequently, stratum 4 decreased (among Phase I respondents) from 410 facilities to 208 facilities.

Because of the significant number of facilities that switched modalities between sample selection and the Phase I interview, the stratification of facilities prior to Phase II sample selection was based on responses to the Phase I interview. The Phase I data base was thought to contain more reliable and more up-to-date information about treatment modality and other characteristics of the facilities.

Table 3-13. Reclassification of facilities-NFR strata designation to Phase I analytic strata (unweighted)

	Hospital inpatient		Residential		Outpatient methadone		Outpatient alcohol		Outpatient other		Combined		Total	
NFR strata	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Hospital Inpatient	194	54.96	42	11.90	0	0	3	0.85	37	10.48	77	21.81	353	100
Residential	1	0.27	320	85.56	2	0.53	0	0	7	1.87	44	11.76	374	100
Outpatient-Methadone	0	0	1	0.26	351	91.41	3	0.78	28	7.29	1	0.26	384	100
Outpatient-Alcohol	0	0	2	0.49	7	1.71	126	30.73	272	66.34	3	0.73	410	100
Outpatient-Other	1	0.26	3	0.77	12	3.08	32	8.23	333	85.60	8	2.06	389	100
Combined	4	1.52	23	8.75	4	1.52	8	3.04	83	31.56	141	53.61	263	100
Unknown	3	1.35	37	16.67	7	3.15	36	16.22	131	59.01	8	3.60	222	100
Total	203	8.48	428	17.87	383	15.99	208	8.68	891	37.20	282	11.77	2,395	100.0

To accurately predict the appropriate migration rates for all facilities in the ADSS sampling frame, the migration rates were weighted using the full sample weights. Section 5.1.3 discusses the calculation of the full sample weights. Table 3-14 shows the weighted migration rates for respondents to the Phase I questionnaire. The figures in the table represent the estimated total number of eligible facilities in the ADSS sampling frame. For example, the Phase I data estimates that about 36 percent of the facilities assigned to NFR stratum 4 would remain in stratum 4, and about 63 percent would have migrated to stratum 5.

A comparison of the unweighted and weighted tables allows one to analyze the size of the facilities that migrated. In general, the weighted migration pattern shows rates substantially lower than the unweighted migration pattern for facilities that remain in stratum 1. For facilities that remained in strata 4 and 6, the weighted rates are higher than the unweighted rates. This means that in general, the small facilities tended to remain in stratum 1, while large facilities in strata 4 and 6 tended to remain in their respective strata.

Table 3-14. Weighted migration pattern for facilities responding to Phase I of ADSS

	Hospital inpatient		Residential		Outpatient methadone		Outpatient alcohol		Outpatient other		Combined		Total	
NFR strata	$\hat{N}$	%	$\hat{N}$	%	$\hat{N}$	%	$\hat{N}$	%	$\hat{N}$	%	$\hat{N}$	%	$\hat{N}$	%
Hospital Inpatient	325.38	47.81	82.55	12.13	0	0	13.15	1.93	116.86	17.17	142.60	20.95	680.54	100.00
Residential	3.24	0.18	1,519.80	84.39	8.55	0.47	0	0	33.00	1.83	236.39	13.13	1,800.98	100.00
Outpatient-Methadone	0	0	1.04	0.23	418.76	91.11	3.82	0.83	34.84	7.58	1.18	0.26	459.64	100.00
Outpatient-Alcohol	0	0	4.41	0.31	8.87	0.62	510.71	35.72	894.79	62.58	11.03	0.77	1,429.81	100.00
Outpatient-Other	9.62	0.23	59.28	1.41	63.39	1.51	416.24	9.88	3,607.00	85.66	55.36	1.31	4,210.89	100.00
Combined	11.14	0.58	165.53	8.66	12.65	0.66	82.52	4.32	429.60	22.47	1,210.10	63.31	1,911.54	100.00
Unknown	29.07	1.53	301.96	15.94	53.36	2.82	329.07	17.37	1,118.00	59.02	62.65	3.31	1,894.11	100.00
Total	378.45	3.06	2,134.57	17.23	565.58	4.57	1,355.51	10.94	6,234.09	50.33	1,719.31	13.88	12,387.51	100.00

$\hat{N}$ = sum of weights, an estimate of the number of eligible facilities.

4. WEIGHTING PROCESS

The estimation process for Phase I involved computing sampling weights to account for unit nonresponse and involved imputation (discussed in Section 5) to account for item nonresponse. Weights were applied to sample data to estimate statistics. The weights for Phase I were processed in several stages in order to accomplish the following objectives:

- Poststratification attempted to adjust for undercoverage of weighted estimates due to the facility measure of size used in sample selection and was used to improve the precision of survey estimates;
- Trimming reduced the impact of extreme weights on the variance and mean squared error of survey estimates;
- Adjustment for nonresponse compensated for unit nonresponse in Phase I and attempted to reduce nonresponse bias due to differences between nonrespondents and respondents; and
- The replication procedure produced replicate weights that are used to compute sampling error estimates.

Full sample and replicate weights were produced and the attached flowchart (Exhibit 4-1) outlines the process for creating full sample weights and replicate weights for Phase I. Details of the procedure used to compute the weights will follow.

To account for item nonresponse, imputed values will be used in place of missing information for items on the Phase I questionnaire. Hot-deck imputation is a technique where missing items are replaced by reported items from other facilities (donors) with similar characteristics. Plans to use a hot-deck technique for the imputation procedure will be discussed.

The Phase I weighting procedures began with updating the weighting variables with the responses given in the Phase I questionnaire. Base weights were computed and poststratified to frame counts. The poststratified weights were trimmed, and then the trimmed weights were adjusted for unit nonresponse in order to attain the full sample weights. Replicate weights were created using the same procedures and adjustments that were used to arrive at the full sample weights. Finite population correction factors were computed to account for sampling without replacement from a finite population. Some guidelines for computing degrees of freedom was given in order to help with the analyses of Phase I data. The following sections explain the weighting procedure in detail.

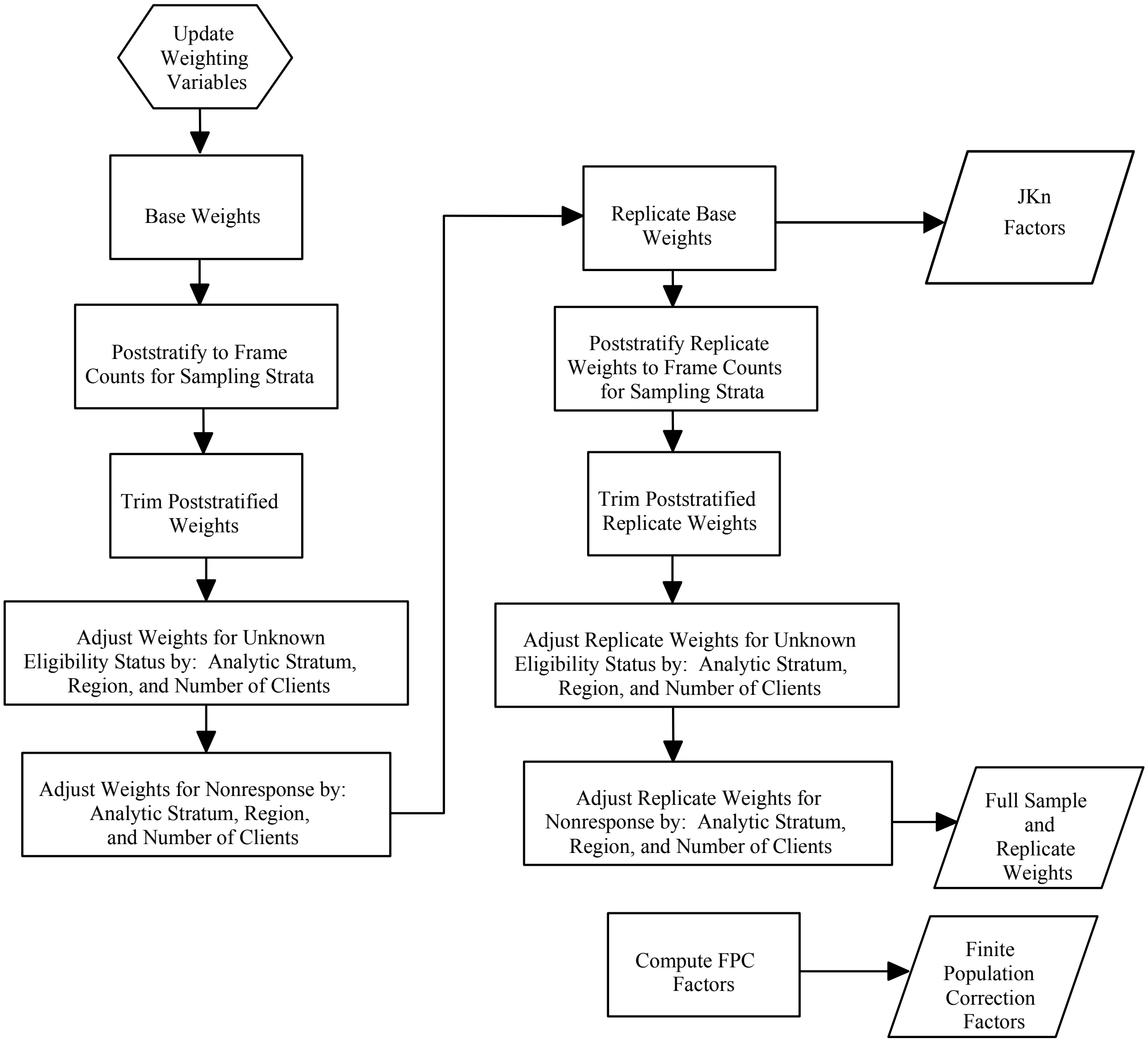
Update Weighting Variables
Base Weights
Poststratify to Frame Counts for Sampling Strata
Trim Poststratified Weights
Adjust Weights for Unknown Eligibility Status by: Analytic Stratum, Region, and Number of Clients
Adjust Weights for Nonresponse by: Analytic Stratum, Region, and Number of Clients
Replicate Base Weights
JKn Factors
Poststratify Replicate Weights to Frame Counts for Sampling Strata
Trim Poststratified Replicate Weights
Adjust Replicate Weights for Unknown Eligibility Status by: Analytic Stratum, Region, and Number of Clients
Adjust Replicate Weights for Nonresponse by: Analytic Stratum, Region, and Number of Clients
Full Sample and Replicate Weights
Compute FPC Factors
Finite Population Correction Factors

Exhibit 4-1. ADSS Cycle 1 Phase I Weighting

4.1 Data Cleaning

Before weights were created, it was necessary to update some variables with Phase I questionnaire data in order to have the most current information available when computing the Phase I weights. The weighting variables were used to construct adjustment cells for adjusting the weights due to unit nonresponse. Weighting variables that were updated include analytic stratum, number of clients, type of ownership, and census region.

There were seven sampling strata in Phase I, one of which contained facilities that could not be placed in the other six strata based on the limited information available in the sampling frame. Phase I respondents were assigned to one of six analytic strata based on their responses to the Phase I questionnaire. If facilities did not respond to the Phase I questionnaire, they were assigned to an analytic stratum based on its screener response. Facilities were assigned to its sampling stratum if they did not respond to the Phase I questionnaire and did not respond to the screener. After reassigning facilities to analytic strata, there were some facilities that were not assigned to one of the six analytic strata. For these cases, analytic stratum were randomly imputed based on the migration rate of sampling stratum 7 facilities that migrated into the six analytic strata.

The number of clients variable was categorized to create nonresponse adjustment cells. Before it was categorized, the number of clients variable was updated using Phase I questionnaire data. For children of administrative units, the number of clients on the Phase I questionnaire was used if reported. If the number of clients was not reported on the Phase I questionnaire, then the average number of clients per child facility within the associated parent facility was imputed. The categorized version of the number of clients was created using four categories: 0-16, 17-40, 41-100, and More than 100 clients.

The type of ownership variable was recoded from its categories on the Phase I questionnaire, to the categories it has on the sampling frame. It was necessary to collapse the state and local government categories together to be consistent with the coding used for the frame variable. The Phase I data, if available, was used to update the frame information on the type of ownership.

The region variable was missing for children of administrative units and another case that was added due to its Phase I response. The region of these facilities was imputed using the state information that was available for these cases.

Response flags were assigned based on the facilities' questionnaire status and screener status. There were four categories of response: respondents, nonrespondents, ineligibles, and unknown eligibility status. Facilities that were not locatable, not available, given maximum number of calls, were among those given unknown eligibility status. A small number of the ineligible facilities were actually identified as a result of the extended interview conducted in Phase II. Table 4-1 shows the distribution of the response status code by analytic stratum.

Table 4-1. Distribution of response status code by analytic stratum

Analytic stratum	Respondents	Nonrespondents	Ineligibles	Unknown eligibility status	Total
1	203	17	240	1	461
2	428	19	173	2	622
3	383	22	65	8	478
4	208	45	185	5	443
5	891	66	259	12	1,228
6	282	23	103	3	411
Total	2,395	192	1,025	31	3,643

4.2 Base Weights

The base weights were computed for 3643 facilities. The number of facilities include 3506 records from the screener sample, 93 children of administrative units, 43 deselected cases that were flagged after the Phase I questionnaire was sent, and one facility added that had branched off from a sample facility. Of the 93 children of administrative units, 2 were children of a deselected unit. Therefore, the 45 (43 + 2) deselected cases were initially given base weights of one to indicate that they only represent themselves. The remaining 91 children of administrative units and the added facility were given the probability of selection associated with their parent facility or the original sample facility in the case of the added facility. The probabilities of selection were adjusted after identifying sample cases that had duplicates on the frame. These probabilities of selection that account for duplication (see Section 3.11 for a discussion on duplicates) were trimmed to one and the base weights were computed as the inverse of the probability of selection. For sampled facilities that are duplicates of other sampled facilities, the base weights for the ineligible records were set to zero.

In order to minimize the degree of overlap between ADSS and NESAT, the Permanent Random Number (PRN) approach was used to assign facilities in the metro certainty PSUs for both ADSS and NESAT. The weights were computed as the inverse of the probabilities of selection in certainty PSUs. In noncertainty PSUs, however, the measure of size was doubled for facilities in the 26 metro noncertainty ADSS PSUs, and were set equal to zero in the 26 metro noncertainty PSUs in NESAT (the ADSS and NESAT noncertainty PSUs in the Westat 100 PSU Master Sample). In addition, the probabilities of selection for facilities in the ADSS nonmetro PSUs were doubled to assure the inclusion of an adequate sample of facilities from rural areas. Since the resulting weights for certainty facilities were equal to one, weights of certainty facilities in the noncertainty PSUs were doubled in order to represent the certainty facilities in the NESAT metro noncertainty PSUs. For the methadone stratum, a more exact adjustment was applied to the weights based on the number of facilities assigned to NESAT and ADSS (see Table 4-2). After this adjustment to the methadone stratum facilities, the sum of the base weights for facilities sampled in the methadone stratum (not including children of administrative units and deselected cases) was equivalent to the frame count.

Table 4-2. Adjustment factors for the methadone stratum

FIELDPSU	*METH_ADJ*
112 or 113	5
119 or 110	3/2
114 or 103	5/4
104 or 111	2
219 or 204	3
209 or 210	3/2
322 or 306	3
331 or 332 or 320	10/3
319 or 314 or 308	4
410 or 408 or 419	10/3
406 or 401	5/3
417 or 402	2
all other PSUs	1

There were 28 administrative units identified by the screener questionnaire. Two of the 28 administrative units requested that one questionnaire be filled out for all its associated children. The 16 child facilities of the two administrative units were coded as ineligible. The eligibility status of the two administrative units was based on the status of the administrative units itself. For the remaining 26 parent facilities, there were 77 child facilities. The 26 parent facilities were coded as ineligible. The probability of selection of the parent facility was given to its associated child facilities. The base weights of the child facilities were computed as the inverse of the selection probability of its associated parent. Their base weights were poststratified and trimmed just as other facilities' weights.

4.3 Poststratification Adjustments

The base weights were poststratified to bring their sum, by PSU type within each sampling stratum, to the frame counts. The poststratification procedure was applied to the Phase I base weights because the measure of size to which probabilities of selection were based, was computed using the number of clients raised to the 0.7th power (as described in Section 3.4). The resulting base weights were relative weights, and were not reliable for estimating totals. To reduce the variance of the estimates, poststratification is often used (Cochrane, 1977). Poststratification adjustment factors were computed for each PSU type, *j*, and sampling stratum combination, *h*, as the ratio of the sampling frame count, N_{jh}, to the sum of the weights, w'_{jhi}, across n'_{jh}, the number of sampled facilities in PSU type *j* and sampling stratum *h* that are not deselects or children of administrative units. The initial facility base weights, w'_{jhi}, were computed strictly for the purpose of computing poststratification factors, and thus excluding children facilities of administrative units, facilities that were deselected but remained in the sample. In addition, adjustments for duplication were not applied for this computation since duplicates existed in the frame and thus contributed to the frame counts. The poststratification adjustment factors were computed as,

$$F_{jh}^{(1)} = \frac{N_{jh}}{\sum_{i=1}^{n'_{jh}} w'_{jhi}} .$$

The poststratification adjustment factors were applied to the facility base weights, w_{jhi}, for the 3,643 facilities (which includes the children facilities of administrative units, and deselects in the sample). The computation of the facility base weights is explained in Sections 4.2 and 4.3.

$$w_{jhi}^{(1)} = F_{jh}^{(1)} w_{jhi} .$$

Note that as a result of this adjustment, some poststratified base weights fall between zero and one.

The frame counts to which the sum of weights were adjusted are given in Table 4-3, along with the adjustment factors.

Table 4-3. Frame counts by sampling stratum and PSU type

	Certainty MSA		Noncertainty MSA		Noncertainty Non-MSA		Total
Sampling stratum	Frame count	Adjustment factor	Frame count	Adjustment factor	Frame count	Adjustment factor	Frame count
Hospital Inpatient	327	1.06	572	1.23	269	0.96	1,168
Residential	716	0.98	1,246	1.25	367	0.94	2,329
Outpatient-Methadone	282	1.00	216	1.00	13	1.00	511
Outpatient-Alcohol	569	1.11	824	1.19	670	1.02	2,063
Outpatient-Other	1,906	0.86	3,038	1.27	1,280	0.97	6,224
Combined	598	1.21	1,416	1.32	561	0.99	2,575
Unknown	936	1.05	1,820	1.21	742	0.98	3,498
Total	5,334		9,132		3,902		18,368

4.4 Trimming

The poststratified base weights were trimmed only if necessary since trimming introduces bias into the survey estimates. Weights were trimmed if they were considered an outlier among other weights in the same analytic stratum. The stratum mean of the weights multiplied by four was used as a general guide to identify potential outliers after extensive diagnostic checks were conducted on the distribution of facility weights within strata. In addition, the weight trimming procedure took into account the fairly high level of facilities that migrated from their sampling stratum. After weights were trimmed, the excess weight from the trimmed weight was distributed to the untrimmed weights to maintain the level of the sum of weights. Table 4-4 shows the maximum weights before and after trimming, and also shows the number of weights that were trimmed within each analytic stratum. The trimmed weights are called $w^{(2)}$ for future reference.

Table 4-4. Maximum weights before and after trimming

Analytic stratum	Maximum weight before trimming	Maximum weight after trimming	Number of trimmed weights
1	13.02	9.03	1
2	44.34	23.02	1
3	16.80	9.03	2
4	49.01	27.66	1
5	72.79	48.24	4
6	45.28	32.41	2

4.5 Nonresponse Adjustment

Nonresponse adjustment was carried out in two phases. The first phase consisted of distributing the weights of records with unknown eligibility status to the weights of records with known eligibility status. To do this, adjustment cells were created in the hierarchy of analytic stratum (six levels), census region (four levels), and categorized number of clients (four levels), creating 96 cells. Type of ownership was considered for nonresponse adjustment cell construction. However, this variable was dropped from consideration because of the amount of missingness on the sampling frame. The type of PSU was also considered initially as a weighting variable, but was a weak predictor of response status during a preliminary investigation using the software CHAID. The software was used to find important predictors of response propensity by creating groups so that the response rate within cells is as constant as possible, and the response rate between cells is as different as possible. The CHAID results were used as a guide to organize the hierarchy of variables for collapsing cells during the nonresponse adjustment procedure.

The nonresponse adjustment cells were collapsed when the adjustment factor was greater than two or the number of eligible units was less than 30. The 96 cells were collapsed to 56 cells based on the collapsing criteria. Cells were not collapsed across strata. The client categories were collapsed within the corresponding level of region. The adjustment factor due to unknown status was computed within each cell as:

$$F_c^{(2)} = \frac{\sum_{i=1}^{n_1} w_{ci}^{(2)} + \sum_{i=1}^{n_2} w_{ci}^{(2)} + \sum_{i=1}^{n_3} w_{ci}^{(2)} + \sum_{i=1}^{n_4} w_{ci}^{(2)}}{\sum_{i=1}^{n_1} w_{ci}^{(2)} + \sum_{i=1}^{n_2} w_{ci}^{(2)} + \sum_{i=1}^{n_3} w_{ci}^{(2)}};$$

where, n_1 = number of respondents, n_2 = number of nonrespondents, n_3 = number of ineligibles, n_4 = number with unknown eligibility status. After the weights of unknown eligibility status were distributed, the weights for records with unknown status were set to zero. Weights associated with respondents, nonrespondents, and ineligibles were adjusted as follows:

$$w_{ci}^{(3)} = F_c^{(2)} w_{ci}^{(2)}.$$

The second phase of nonresponse adjustment consisted of distributing the weights of the nonrespondents to the weights of the respondents. The resulting cells from the previous adjustment procedure were used and further collapsed when the adjustment factor was greater than two or the number of respondents was less than 30. The 58 cells from the first phase were collapsed to 43 cells based on the collapsing criteria. The adjustment factor due to nonresponse was computed within each cell as:

$$F_c^{(3)} = \frac{\sum_{i=1}^{n_1} w_{ci}^{(3)} + \sum_{i=1}^{n_2} w_{ci}^{(3)}}{\sum_{i=1}^{n_1} w_{ci}^{(3)}}.$$

The weights of nonrespondents were set to zero. The weights of respondents were adjusted due to nonresponse by applying the nonresponse adjustment factor to the weights that were adjusted for unknown eligibility status:

$$w_{ci}^{(4)} = F_c^{(3)} w_{ci}^{(3)}.$$

After nonresponse adjustment, 29 facility weights remained between zero and one as a result of poststratification (see Section 4.3). These weights were inflated to one. The result is the full sample weight. Table 4-5 shows the sum of the full sample weights and distribution of weights for each analytic stratum for responding facilities. Table 4-6 gives the sum of weights after each stage of weighting.

Table 4-5. Full sample weights

Analytic stratum	Sum of weights	Minimum	10th percentile	Median	90th percentile	Maximum
1	378.45	1.03	1.14	1.40	3.05	10.71
2	2,134.53	1.00	1.31	4.02	9.97	29.56
3	565.58	1.07	1.09	1.09	1.82	9.88
4	1,355.50	1.16	1.44	4.31	12.79	38.15
5	6,234.08	1.00	1.30	4.72	13.14	60.28
6	1,791.32	1.00	1.33	3.60	14.14	43.34

Table 4-6. Sum of weights after each stage of weighting

Weights	Number of records	Sum of weights
Base Weights	3,643	16,802.66
Poststratified Base Weights	3,643	18,432.84
Final Poststratified Weights	3,643	18,164.49
Trimmed Weights	3,643	18,164.49
Unknown Status Weights	3,643	18,164.49
Nonresponse Adjusted Weights	3,643	18,164.49
Final Full Sample Weight	3,643	18,165.74

4.6 Variance Estimation

A class of techniques called *replication methods* provides a general method of estimating variances for the types of complex sample designs and weighting procedures usually encountered in practice (Wolter, 1985). The basic idea behind the replication approach is to select subsamples repeatedly from the whole sample, to calculate the statistic of interest for each of these subsamples, and then to use the variability among these subsample or replicate statistics to estimate the variance of the full sample statistics. There are different ways of creating subsamples from the full sample. The subsamples are called *replicates* and the statistics calculated from these replicates are called *replicate estimates*.

Replicate weights were created using a variation of the stratified jackknife procedure. The sample was divided into 12 variance strata based on the sampling stratum and whether the facility was selected with certainty. Two hundred replicates were identified by forming n_h random groups within each variance stratum, *h*, so that $\sum n_h = 200$. The n_h were calculated for each stratum by proportionally allocating the 200 replicates according to the number of facilities within each variance stratum.[1]

The weights for each replicate were formed by setting the full sample base weights for facilities in random group *g* within variance stratum *h* to zero, adjusting other weights within the stratum to account for the 'dropped' units, and leaving the weights for all remaining observations unchanged. This procedure was performed for all replicates and variance strata, forming 200 sets of replicate weights.

[1] Two variance strata, those containing certainty facilities in sampling stratum 5 (Outpatient – All Other) and certainty facilities in stratum 7 (Unknown), were combined with other variance strata since the number of facilities in these strata was too few to allow more than one replicate. Since variance strata must have more than one replicate when using the JKn procedure, certainty facilities in stratum 5 were combined with those in stratum 4 (Outpatient – Almost Exclusively Alcohol); certainty facilities in stratum 7 were combined with those in stratum 6 (Combined).

The formula for estimating the variance is computed as,

$$v(\hat{\theta}) = \sum (\hat{\theta}_g - \hat{\theta})^2 .$$

where, $\hat{\theta}$ is the parameter estimate for θ, and $\hat{\theta}_g$ is the parameter estimate for replicate g using the associated replicate weights.

As an alternative to the replication method, the Taylor series method can be used to approximate variances under complex sample designs. Computer software packages have been developed to analyze data from complex samples using the replication and/or Taylor series methods. Please refer to the ADSS data codebook for more information about software packages that offer replication and Taylor's series methods, specifically WesVar (WesVar Complex Samples 3.0[2]), SUDAAN[3] (Software for the Statistical Analysis of Correlated Data), and Stata[4]. Any of the three packages can be used in the analysis of the ADSS data. The information in the ADSS data codebook includes a discussion on software capabilities and is presented to help users select the software most appropriate for their analysis.

4.6.1 Adjustments to the Replicated Weights

The same adjustments made to the base weights were done to the replicate weights. These adjustments included poststratification, trimming, and nonresponse adjustment. For each replicate, poststratification adjustment factors were computed for each PSU type within each sampling stratum as the ratio of the sampling frame count to the sum of weights, where the sum of weights was across the 3,506 records, and the weights did not account for duplication on the frame. The resulting poststratification adjustment factors were applied to each replicate weight for each of the 3,643 sample cases.

[2] WesVar is developed by Westat (www.westat.com) and distributed by SPSS, Inc. (www.spss.com).

[3] SUDAAN is developed and sold by the Research Triangle Institute (www.rti.org).

[4] Stata is a registered trademark of Stata Corporation (www.stata.com).

For trimming the replicate weights, the process was automated. The difference in the procedures was in identifying the largest acceptable weight within the replicate and within the same analytic stratum. The largest weight to which larger weights were trimmed was based on the ratio of the largest acceptable poststratified base weight to the mean of the poststratified base weights. This ratio was applied to the mean poststratified replicate weight to determine the maximum acceptable poststratified replicate weight for each analytic stratum.

For nonresponse adjustment, the cells from the full sample 'unknown status' adjustments were used initially when adjustment factors for unknown eligibility status were computed for the replicate weights. Further collapsing of cells was based on the criteria of maximum adjustment factor = 2 and a minimum number with known eligibility status of 25. If a cell had a violation in at least one replicate, collapsing occurred for all replicates. The new cells were used for input cells for nonresponse adjustment of the replicate weights. For nonresponse adjustment relating to the replicate weights, collapsing occurred when the adjustment factor was greater than 2 or the minimum number of respondents was less than 25. Again, if a cell had a violation in at least one replicate, collapsing occurred for all replicates.

After all adjustments were made, a small number of replicate weights were less than one for a small number of sample cases due to the poststratification adjustment. The weights for these cases were set to one. No full sample weights were less than one.

4.6.2 Finite Population Correction and JKn Factors

The Finite Population Correction (fpc) factors were computed since sampling was done without replacement from a finite population, and the sampling rate was high enough that the factors could not be ignored. The fpc factors for each replicate were computed based on the number of facilities on the frame and the number of facilities sampled excluding nonrespondents. They are calculated as $f_g = (N_h - n_h) / N_h$, where *h* is the variance stratum associated with replicate *g*, N_h is the frame count for stratum *h*, and n_h is the number of sample cases excluding nonrespondents.

Since Phase I sampling was conducted with unequal probabilities of selection, the above mentioned formula for the fpc factors is not strictly applicable. Currently, this issue is an on-going survey research item. There are alternative points of view about computing and applying fpc factors. First, the resulting fpc factors can be used as an approximate variance reduction tool to account for sampling from a finite population, only if the count of facilities in the sample approximately reflects the contribution to the sampling strata in terms of measure of size. This seems conceptually intuitive, however, there is no theoretical justification for using count-based fpc factors, and for computing measure-of-size-based fpc factors. The count-based fpc factors, as given in the formula for f_g above, were compared to the analogous measure-of-size-based fpc factors. In general, the count-based fpc factors were close to or slightly larger than the corresponding measure-of-size-based fpc factors. Given that the count-based factors will yield slightly more conservative variance estimates, the count-based fpc factors are recommended. The two sets of fpc factors for Phase I replicates are shown below:

Count-based fpc factors:

```
0.05, 0.06, 0.06, 0.05, 0.06, 0.06, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,
0.62, 0.61, 0.62, 0.61, 0.62, 0.62, 0.62, 0.62, 0.62, 0.62, 0.61, 0.61, 0.61,
0.61, 0.61, 0.61, 0.61, 0.61, 0.61, 0.61, 0.61, 0.00, 0.00, 0.80, 0.80, 0.80,
0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80,
0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.05,
0.05, 0.04, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05, 0.05,
0.05, 0.05, 0.33, 0.32, 0.33, 0.33, 0.33, 0.33, 0.33, 0.32, 0.33, 0.07, 0.07,
0.06, 0.07, 0.07, 0.07, 0.07, 0.08, 0.78, 0.78, 0.78, 0.78, 0.78, 0.78, 0.78,
0.78, 0.78, 0.78, 0.78, 0.78, 0.78, 0.78, 0.78, 0.78, 0.78, 0.78, 0.78, 0.78,
0.78, 0.78, 0.78, 0.78, 0.78, 0.78, 0.78, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93,
0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93,
0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.93, 0.02, 0.00, 0.02, 0.02,
0.88, 0.87, 0.88, 0.87, 0.88, 0.87, 0.88, 0.88, 0.87, 0.87, 0.88, 0.88, 0.87,
0.88, 0.87, 0.88, 0.88, 0.88, 0.87, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88,
0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88,
0.88, 0.88, 0.88, 0.88, 0.88.
```

Measure-of-size-based fpc factors:

```
0.07, 0.07, 0.07, 0.07, 0.07, 0.07, 0.04, 0.07, 0.07, 0.06, 0.07, 0.07, 0.07,
0.43, 0.43, 0.44, 0.45, 0.45, 0.44, 0.43, 0.42, 0.43, 0.43, 0.42, 0.42, 0.43,
0.44, 0.43, 0.42, 0.43, 0.43, 0.43, 0.43, 0.42, 0.00, 0.00, 0.69, 0.68, 0.69,
0.69, 0.68, 0.68, 0.68, 0.69, 0.69, 0.69, 0.69, 0.68, 0.68, 0.69, 0.69, 0.69,
0.69, 0.69, 0.68, 0.69, 0.69, 0.70, 0.69, 0.68, 0.68, 0.68, 0.69, 0.69, 0.06,
0.06, 0.04, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.06, 0.05, 0.06,
0.06, 0.06, 0.36, 0.36, 0.36, 0.37, 0.36, 0.37, 0.35, 0.34, 0.36, 0.06, 0.06,
0.05, 0.05, 0.06, 0.06, 0.06, 0.07, 0.65, 0.65, 0.65, 0.65, 0.65, 0.65, 0.66,
0.65, 0.65, 0.65, 0.66, 0.65, 0.65, 0.65, 0.65, 0.65, 0.65, 0.65, 0.65, 0.65,
0.66, 0.65, 0.66, 0.65, 0.66, 0.65, 0.65, 0.87, 0.88, 0.88, 0.88, 0.88, 0.88,
0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.87,
0.88, 0.88, 0.88, 0.87, 0.88, 0.88, 0.88, 0.88, 0.88, 0.02, 0.00, 0.02, 0.02,
0.79, 0.79, 0.79, 0.79, 0.79, 0.79, 0.79, 0.79, 0.79, 0.79, 0.79, 0.79, 0.79,
0.79, 0.79, 0.79, 0.79, 0.79, 0.79, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88,
0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88, 0.88,
0.88, 0.88, 0.88, 0.88, 0.88.
```

The fpc factors are small for some replicates due to being aligned with variance strata associated with certainty units. Variance contributions were allowed from variance strata that were constructed for certainty facilities by assuming that nonresponse occurred at random within the variance strata.

The second point of view is to infer results from ADSS analyses to a superpopulation, that is, to assume that the finite population of eligible facilities for ADSS comes from a superpopulation of facilities. In that case, applying fpc factors would be inferentially incorrect. The idea of superpopulation applies well to finite populations for which members go through moderate to substantial changes overtime. The ADSS target population is a rather stable population of facilities, and thus, this approach is not recommended for ADSS.

The JKn factors were calculated as $k_g = (n_h - 1) / n_h$, for each replicate *g*, where *h* is the variance stratum associated with replicate *g*, and n_h is the number of random groups, or replicates.

The fpc factors, f_g, and the JKn factors, k_g, are used in the variance computations as follows:

$$v(\hat{\theta}) = \sum f_g k_g (\hat{\theta}_g - \hat{\theta})^2 ;$$

where, $\hat{\theta}$ is the parameter estimate for θ, and $\hat{\theta}_g$ is the parameter estimate for replicate *g*.

For variables that contain imputed values, to simplify the computation of variances in the presence of imputation error, the approach recommended for ADSS is to incorporate the imputation error variance by using a variance inflation factor (VIF) (discussed in Section 5.11). The variance inflation factor can be multiplied by the variance (computed by treating imputed values as if they were observed) after the calculation of the jackknife variances.

4.6.3 Degrees of Freedom

The Degrees of Freedom (DF) associated with the variance estimator could affect the outcomes of statistical tests for analyses on subsets of the total ADSS sample. The approximate number of degrees of freedom can be used to approximate the actual degrees of freedom. If the approximate number of degrees of freedom is greater than about 30, then the impact of the degrees of freedom on the analyses may be considered negligible. The approximate degrees of freedom for the JKn variance estimator used for ADSS Phase I can be calculated as the number of active replicates less the number of active variance strata. For an analysis using the entire sample, all 200 replicates and 12 variance strata are active; that is, each variance unit that is aligned with a replicate includes sampled cases. Therefore the approximate degrees of freedom is 200 – 12, or 188. The approximate degrees of freedom (by analytic stratum) are in Table 4-7 below.

Table 4-7. Approximate degrees of freedom (by analytic stratum)

Analytic stratum	Approximate DF
1	44
2	95
3	52
4	92
5	147
6	96

Tabular or regression analyses performed at the analytic stratum level can ignore the effects of degrees of freedom on the inferences. For any subset of the sample that analyses are run, we encourage the calculation of the approximate degrees of freedom. Using the Phase I weighting files, one can compute the approximate degrees of freedom by counting the number of active replicates associated with the subset. For instance, suppose an analysis will be done on the subset analytic stratum 1 and region 1. To count the number of active replicates, produce a frequency on the variance unit and variance stratum variables for the subset. The number of active replicates will equal the number of unique combinations of variance strata and variance units remaining on the data. The approximate degrees of freedom may be calculated by then subtracting the number of unique variance strata remaining on the data. Recall that the variance stratum and variance unit together identify the replicate to which the sample unit is aligned.

In general, the actual effective degrees of freedom are less than the approximate degrees of freedom. This is due to the complex nature of the sample design, specifically the level of between-facility variance and the relative size of the stratum in terms of the domain being analyzed. Therefore, the data analyst should use caution when making inferences on small domains.

5. IMPUTATION PROCESS

There are two types of nonresponse in ADSS, facility-level (or unit) nonresponse and questionnaire item nonresponse. Unit nonresponse occurs when few or no survey items are obtained from a sampled facility. This arises for several reasons including when the facility refused to respond, was unable to participate, or could not be located. Sample weighting, as discussed in Section 4, was used to compensate for unit nonresponse.

Item nonresponse occurs when some items of the questionnaire are left blank due to inadvertent omissions, refusals, lack of knowledge, or edit failures. Item nonresponse in ADSS is handled for some key items through imputation. For a list of the imputation items, see Appendix B. Imputation is used to reduce nonresponse bias in survey estimates, simplify analyses, and make results consistent across analyses. In general, imputation is the last step in the estimation process and is used to fill-in data for missing values among the unit respondents.

This section provides details of the imputation process. The general process involved several tasks in the following order:

- Edits;
- Logical imputation;
- Imputation using external sources;
- Statistical imputation to fill in the remainder of missing values; and
- Measuring imputation error variance.

5.1 Edits

The ADSS Phase I pre-imputation file served as input for the imputation process. Pre-imputation edit checks were processed for each item so that inconsistencies in the data could be identified, for instance, individual parts not summing to a total or values not copied correctly to another item. Data records that contained inconsistencies were not used for imputation model building or as donors of imputed values for relevant data items.

Table 5-1 shows the frequency of data inconsistencies that remain. The inconsistency flags are available on the ADSS Phase I imputation file. Each flag identifies a record that contains an inconsistency within a group of imputation variables. The inconsistencies did not result from the imputations, but resulted from respondent error or entry errors.

Table 5-1. Data inconsistencies in the ADSS pre-imputation Phase I file

Flag name	Items relating to the flag	Number of records with inconsistencies
A9FLG	A9 matrix	15
B1FLG	B1 matrix	27
B2FLG	B2 matrix	31
B3FLG	B3	0
B4FLG	B4	0
B12FLG	B12A, B12B	3
C2FLG	C2 Column 1	34
C4AFLG	C4A, C4AFLG, C2F1	3
C4BFLG	C4B, C4BFLG, C2F1	16
D4FLG	D1, D4	0
D8FLG	D7, D8 items	11
D12FLG	D7, D12 items	4
D13FLG	D7, D12C, D13 items	13
D15FLG	D14, D15 items	14
D16FLG	D14, D16 items	66

5.2 Logical Imputation

Logical imputation is a procedure used to impute for missing values on the data file, where the true (but not reported) values can be deduced using other data that the facility has reported. Missing values that can be filled-in with logical imputations are those for which, based on the reported relevant data, only one specific value can be assigned to the missing case. For example, in a grid with 15 males, missing females, and a total of 15 clients, the number of females should be coded as 0.

Other situations where logical imputation was applied included filling-in data that should have been copied from another item and also converting percentages to dollar values. In addition, imputation was initially done on dollar values, then logically converted to percentages, where the percentage was missing. For example, some financial questions requested that either a percentage or a dollar amount is needed. When only the percentages were reported, then the percentages were converted to dollar values based on the grand total for that particular item. Sometimes it was necessary to round the dollar values that were converted from percentages, so that the sum of the dollar values added to the total reported. If the converted dollar values were not consistent with the reported percentage and the reported percentages did not add to 100, then the dollar values were rounded to be consistent with the reported percentages, and the data record was left with the percentage inconsistencies as they were reported.

5.3 External Sources

The use of data from external sources for Phase I imputation follows efforts to logically impute the data and preceded the use of statistical methods. The general guideline was to use information from the same facility from another source, regardless of the reporting period, rather than use data from another facility or model (imputation). There were three external sources considered, 1) NESAT, 2) Uniform Facility Data Set (UFDS), and 3) Phase II. It was decided not to use data from NESAT, primarily since the sampling unit (service delivery units) was different from the Phase I sampling unit (facility). The UFDS and ADSS Phase II data were used to fill-in missing data. The ordering of the process consisted of 1996 UFDS, then ADSS Phase II data, and then 1997 UFDS.

5.3.1 1996 UFDS

The 1996 UFDS asks for the number of clients (by type of care and age, race and sex), total number of admissions, and total revenue (broken out by subcategories). The UFDS does not ask for staffing or cost data.

First, the UFDS file and the ADSS file were matched. Then ADSS records with missing values among the matched cases were printed out. For the ADSS records with missing ADSS values, the corresponding UFDS data were listed. The listings were reviewed and then the UFDS data were used wherever feasible. Acceptable UFDS value ranges were computed to use as a rule of accepting UFDS values as acceptable substitutes for ADSS responses. Imputed UFDS values that existed in the 1996 UFDS file were not used. Appendix C shows the imputation rates for each imputation item by source of imputation. The 1996 UFDS data was used much more than the Phase II and 1997 UFDS data. The largest imputation rate attributable to UFDS is 3.79 percent for hospital inpatient revenues (3.5 percent specifically for 1996 UFDS).

5.3.2 ADSS Phase II Data

The use of Phase II data was limited to the following items: admissions (C2 Column 1), revenues (D7), and costs (D14). No missing values were filled-in using Phase II data for staffing or client items. Table 5-2 shows the result of comparing Phase I and Phase II questionnaires.

Table 5-2. Comparison of Phase I and Phase II data items

Item name	Phase I item number	Phase I	Phase II
Staffing Matrix	A9	As of October 1, 1996	Updates Phase I response with the current data
Active clients in treatment	B1	As of October 1, 1996	Current information requested
Other clients questions	D1, D4, B2, B3, B4, B12A, B12B		Not asked
Admissions	C2 Column 1	12 month period, could vary	Most recent 12 month period
Other admissions	C4		Not asked
Revenues	D7	Asks for 12 month time period	Most recent 12 month period
Costs	D14	12 month period	Most recent 12 month period
Other financial questions	D8, D12, D13, D15, D16		Not asked

Phase II data was used wherever possible, with the exception of facilities that changed organizational structure (i.e., they offer different types of care). The potential Phase II cost values were also compared to the corresponding Phase I reported value for revenues. Since Phase II data collection was not complete, an incomplete but cleaned Phase II file was used, created on September 25, 1998. The imputation rate attributable to Phase II data was very low; 0.29 percent (hospital inpatient costs) was the largest rate.

5.3.3 1997 UFDS

After the 1996 UFDS and Phase II data were used to fill in missing values on the ADSS Phase I data, the 1997 UFDS data were matched to any remaining records with missing values for the number of clients (by type of care, age, race, and sex), total number of admissions, and total revenue (by subcategory). Recall that UFDS does not ask for staffing or cost data.

The records with missing values for selected fields on the ADSS Phase I data were matched to the UFDS data for 1997. The appropriate UFDS data were listed and examined to determine whether they should be used to fill in the missing data. Since the data from the 1997 UFDS pertains to a different time period than the ADSS data, the facility was checked to be sure that it hadn't changed substantially over time. In particular, of interest was ownership, as this may have affected the revenue and cost information, and the treatment offered. The 1997 UFDS data filled-in 4 (0.2 percent) missing values of D1 (managed care contract?), 9 (0.4 percent) for D4 (percent covered through managed care), and 16 (0.7 percent) for total revenues. It was also used to impute for source of revenues (D8 items).

5.4 Statistical Imputation Procedures

5.4.1 Blocks of Items

The imputation items (Appendix B) in ADSS were arranged into blocks of items where each block had a grand total and the sub-items relating to the grand total. Table 5-3 gives a description of the imputation variables as they were organized into blocks of items. Items were organized into five blocks of items corresponding to the following five groups of variables: clients, admissions, staffing, revenues, and costs. Please refer to Table 5-3 when reading the remainder of the section.

Table 5-3. Blocks of items and items to impute

Block of items	Items to impute
Clients	B1 Matrix, B2 Matrix, B3, B4, B12, D4
Admissions	C2 Column 1, C4A, C4ANUM, C4B, C4BNUM
Staffing	A9 Matrix
Revenues	D7, D12, D13
Costs	D14, D15, D16

The imputation rates for each of the imputation items are presented in Appendix C. A summary of the imputation rates by block of items is shown in Table 5-4.

Table 5-4. Ranges of imputation rates (%) by block of imputation items

		Blocks of items			
Type of care	Clients	Admissions	Staffing	Revenues	Costs
Hospital inpatient	0.3-0.9	1.2		21.0	24.5
Residential	0-0.2	0.7		5.4	5.0
Outpatient	0-0.2	0.2		9.9	11.3
Methadone	0-0.2	1.0		10.1	11.7
Non-methadone	0-0.1	0.1		10.5	11.9
Other					6.7
Grand total	0	0.7	0.2	9.2	10.4
Range for other items	0-1	0.9-12.3	0.1-0.2		7.8-11.2

5.4.2 Background of Imputation Methods

To fill-in the remaining missingness after logical imputations and using external sources, care was taken to preserve multivariate relationships of the observed data. Some imputation methods that are deterministic, for instance imputing the mean within classes or using the predicted value of a regression model, will distort distributions. The best methods to consider are generally those that are stochastic in nature and attempt to maintain the joint distributions between variables. Methods such as random within classes hot-deck procedures (Kalton and Kish, 1984) and random regression (Montaquila and Ponikowski, 1995), improve on maintaining the distribution of the variables relative to the deterministic approaches. There are other ways to attempt to reduce the attenuation of the joint distributions, of note is multiple imputation (Rubin, 1987). Multiple imputation was not used for ADSS since software was not readily available, and attempts were made to be consistent with DSRS imputation. Instead of multiple imputation, single imputations were created through a combination of regression and hot-deck (random within classes. A more detailed discussion of the hot-deck and Bayesian approaches is found in Judkins (1997).

5.4.3 Imputation Procedures for ADSS

The following lists the types of items and the procedures that were used to impute for missing data.

1. Grand totals (total admissions, total full-time staff, total revenues, total costs) -- The random regression imputation procedure (see Section 5.4.5) was used. However, if a sub-item within the block was reported, then the random within-cells hot-deck was used to select a donor, and the donor's ratio of the grand total to the sub-item value was multiplied by the donee's sub-item value, to arrive at the imputed grand total.

2. Type of care items

   - Costs -- Single Modality providers.

   The random within-cells hot-deck proportional allocation method (see Section 5.4.6) was used to impute for the type of care items in order to break out the costs into those costs attributable to the type of modality offered, and those costs not attributable by type of modality.

   - Other type of care items (admissions, revenues) – Single Modality providers;

   - The true value for the single type of care offered was logically deduced from the observed or imputed grand total;

   - Multi Modality providers; and

   - A regression procedure (see Section 5.4.7) was used to maintain the relationships of the proportions of clients, admissions, revenues and costs.

3. Other items

The random within-cells hot-deck proportional allocation method was used to impute for client items, and items not asking for type of care level data.

The general flow of the imputation procedure involved the random regression procedure to impute for a block's grand total, then using the imputation procedures for single modality providers and for the items not associated by type of care to impute the remaining items within the block. Then the next block of items was addressed. The order of the blocks of items started with the block with the least amount of missingness (clients), and ended with the block of items with the most missingness (costs). Since total clients have complete data, the process started by imputing for the clients block using random-within-cells hot-deck imputation. For each record that had at least one missing value within the client's block, a donor was selected so that it could be used to fill-in data for the rest of the client block.

The imputation procedure for multi-modality providers was not implemented until imputations were completed for single-modality facilities. Finally, missing percentages for financial items were logically deduced from observed and imputed dollar values.

Imputations were done separately for the following groups defined by the variable TYPECARE.

TYPECARE = 1, if the facility offers hospital inpatient treatment only.
= 2, if the facility offers residential treatment only.
= 3, if the facility offers outpatient methadone treatment only.
= 4, if the facility offers outpatient nonmethadone treatment only.
= 5, if the facility offers at least two of hospital inpatient, residential, or outpatient.
= 6, if the facility offers both methadone and nonmethadone treatment.

5.4.4 Modeled Data and Eligible Donors

Several records were excluded from the model building process or from being a donor. Table 5-5a presents a list of reasons for excluding cases, using total revenues as an illustration of the pattern of exclusion. There were exclusions made prior to imputing for items other than total revenues. Prior to administering each step of the statistical methods, the following types of records were excluded from the imputation step:

1. Several records were discovered as having outlier values or suspicious relationships between key volume and financial variables. These outliers were excluded from imputation steps relating to the financial items. To begin the check of Phase I items identified as outliers, the original paper surveys were reviewed to ensure the correct values were transferred to the electronic file. Two separate conditions were used to determine outlier responses needing review. Either the response itself or the ratio of two associated responses had to lie more than 2.5 standard deviations from a category's mean. Ratios used in this case included admissions to discharges, patient days based on admissions to patient days based on point prevalence count, point prevalence count to staff, revenue to admissions, revenues to patient days, cost to admissions, and costs to patient days. Incorrect data transfers and cases where margin comments made it clear that some other value more clearly reflected the intended reply of the responder were corrected at this time and treated as original responses.

   Following paper review and the corrections based on it, 200 facilities remained with unresolved outlier responses. Of these 200 facilities, 40 had outliers on the basis of their volume items- admissions, discharges, and length of stay – and another 160 had outliers on the basis of both volume and finance items – staffing, revenue, and cost. The 40 facilities with volume based outliers were called back to confirm or clarify responses. Changes because of such callbacks replaced original responses in the Phase I data file and were not considered outliers. Table 5-5a shows the number of facility outliers of this type (Type 1) by TYPECARE that were excluded from total revenue imputation;

2. Some records had inconsistencies between the type of care offered variable and the items relating to type of care. These records were excluded from modeling and excluded as donors. Table 5-5a shows the number of facility outliers of this type (Type 2) by TYPECARE that were excluded from total revenue imputation;

3. Some records had inconsistencies in their data vector, as discussed in Section 5.1. For instance, individual parts not summing to a total was considered a data inconsistency. The records with data inconsistencies relating to the variables used in the modeling process (or hot-deck process) for the specific imputation step, were excluded from the process. Table 5-5a shows the number of facility inconsistencies of this type (Type 3) by TYPECARE that were excluded from total revenue imputation; and

4. Some facilities reported for a greater entity since they could not breakdown their response to the site. These types of facilities are referred to as multi-site reporters. Multi-site reporters can be identified using the flag MULTSITE in the imputation file. By identifying multi-site reporters through duplicate values reported amongst financial questions, through margin notes on the questionnaire, and by two other financial items (D10BOX and D18BOX), these cases were excluded from the imputation process for financial items. In addition, missing financial items associated with multi-site reporters remained missing. Table 5-5a shows the records associated with multi-site reporting (referred to as Type 4) that were excluded from the model building process for total revenue imputation.

Table 5-5a. Reasons for excluding records from model building - total revenues imputation

	TYPECARE						
Reason type	1	2	3	4	5	6	Total
1 only	24	14	11	36	47	14	146
2 only	2	7	4	18	3	1	35
3 only	4	3	15	23	10	2	57
4 only	2	7	6	31	0	0	46
1, 3	2	0	1	0	1	0	4
1, 4	1	1	0	4	0	1	7
1, 2, 4	0	0	0	1	0	0	1
1, 3, 4	0	1	0	0	0	1	2
2, 3	1	0	2	0	0	0	3
2, 4	0	1	0	4	0	0	5
Total excluded	36	34	39	118	61	19	307
Total records	203	428	324	1,083	282	75	2,395

Table 5-5b shows the percentage of non-missing cases that were excluded in the imputation process. The results are shown by TYPECARE categories for each separate imputation task. Non-missing cases are those facilities that do not have a missing value for at least one item in the block of items being imputed.

The exclusion rates were lowest among the non-financial items. The higher exclusion rates among the financial items are due mostly to reason type 1. (As illustrated for total revenues imputation in Table 5-5a). The remaining records used for imputation were considered the 'best' group to base the imputations.

For financial items, most facilities were able to report revenues and costs relating to substance abuse only, however, others could not. Therefore, prior to using statistical methods for imputation, the financial item values were transformed into values representing substance abuse only, using D10BOX and D10PC for revenues, and D18BOX and D18PC for costs.

Table 5-5b. Percentage of non-missing cases that were excluded from the imputation process, by TYPECARE and task

	TYPECARE					
Task	1	2	3	4	5	6
Hot-deck-clients block	4.0%	3.8%	5.1%	3.4%	5.9%	9.9%
Regression-total admissions	4.0%	2.6%	6.8%	3.8%	6.2%	6.7%
Hot-deck-items C4A, C4B	1.5%	2.2%	3.1%	2.7%	1.9%	2.7%
Hot-deck-admissions block	1.5%	1.9%	5.6%	3.6%	3.7%	1.4%
Hot-tdeck-C4ANUM,C4BNUM	1.3%	1.8%	3.4%	2.8%	2.2%	3.1%
Regression-total FT staff	4.3%	2.4%	6.9%	4.5%	6.5%	6.9%
Hot-deck-staffing block	1.0%	0.2%	0.3%	0.8%	0.7%	0.0%
Regression-total revenues	20.8%	7.9%	12.9%	11.6%	24.0%	25.7%
Regression-total cost	22.6%	8.0%	13.1%	11.3%	25.1%	26.4%
Hot-deck-cost block	17.9%	8.5%	9.6%	13.8%	21.0%	28.0%

Multimodality Imputation Regression	Exclusion rate
Hospital inpatient admissions	5.8%
Residential admissions	6.0%
Outpatient admissions	5.9%
Methadone admissions	10.0%
Non-methadone admissions	10.0%
Hospital inpatient revenues	32.5%
Residential revenues	22.8%
Outpatient revenues	27.3%
Methadone revenues	25.9%
Non-methadone revenues	25.9%
Hospital inpatient costs	36.6%
Residential costs	24.2%
Outpatient costs	30.0%
Methadone costs	32.1%
Non-methadone costs	32.1%

5.4.5 Random Regression Method for Grand Totals

One disadvantage of the hot-deck is that some predictive power is lost by categorizing continuous auxiliary variables. This was improved upon by using random regression imputation for the block's grand total. The random regression approach attempts to preserve relationships between blocks of items. For each grand total item (not including total clients), a stepwise regression by a six level type of care variable was done. Using the significant independent variables, the model was fit within each of the six types of care categories. Outliers, besides those identified in Section 5.4.4, were analytically identified and removed to protect against a small number of observations influencing the parameters of the model. The rule was to exclude cases where the absolute value of the standardized residual was greater than 3.5. After excluding the outlying residuals, the model was refit and residuals examined for constant variance and for normality. The next step was to add a random residual to the predicted values for nonrespondents. The procedure to add error depended on review of the distributions of residuals. The residuals did not conform to constant variance and normality; therefore log transformations for the values of the continuous variables were used. The resulting residual analysis using transformed variables showed constant variance and approximately normal distribution of residuals.

Finally, the imputed value was computed by adding a random draw from the normal distribution to the predicted value, where the random draw was based on the mean and variance of the residuals. The imputed values were used in the modeling process for the next imputation item. The order of imputation was defined by starting with the item with the least amount of missingness, and ending with the item with the most missing values. For the staffing block, respondents could report the number of staff in full-time equivalent units. This occurred about 100 times. The full-time equivalent responses were treated as missing values when imputing for total FT staff so that the full vector of data in terms of full time staff could be used as a predictor variable for revenues and cost grand totals.

Prior to each regression imputation step for grand totals, correlations and stepwise regressions were run for each of the six types of care in order to identify the significant independent variables. Dummy variables for regression models were created from the categorical variables: type of ownership, type of PSU, and census region.

5.4.6 Hot-deck Proportional Allocation

Once a grand total for a block was imputed, it triggered the use of hot-deck imputation proportional allocation, which uses a donor's proportions to fill-in the rest of the items (for client items, single modality providers and costs items, or items not associated by type of care) within the block related to the imputed grand total. The imputed or reported values for the grand total were categorized so that they could be used in forming the cells for the hot-deck. The hot-deck employs a fully interactive model. However, it can handle only a small set of predictor variables or else the cells get too small. Basically, the number of donors within each cell needs to be greater than the number of donees. If not, then the software looks across to another cell by crossing the boundary defined by a soft boundary variable.

With the hot-deck method, multiple variables within the same block of items were imputed using the same donor. This approach was used to help reduce the bias in correlations and to maintain univariate distributions. The same donor was used to impute for each of the individual parts within the block of items by using proportions and ratios within the donor's data record. This procedure attempted to maintain relationships between individual parts within the block of items and also protected against sums over individual items being greater than the total reported for the set of individual items. Donors were put back into the donor pool for the next block of items in order to generate randomness across blocks of items within the same record.

Most of the skip patterns were not a problem since the trigger item (for the skip pattern) was almost always completed (for instance, question B1D1, 'is residential care offered?' is never missing). For a small number of trigger items that have missing values, the trigger item was imputed first, then another donor was used to fill-in the remaining items for the block. The trigger items were included in the list of hard boundary variables to which imputation cells were formed. In addition to the trigger items, the other hard boundary was the six-level type of care grouping.

The categorized versions of the grand totals were used as soft boundaries for the hot-deck imputation procedure (see Appendix D for a description of how the categories were constructed). For instance, for the categorized number of clients, five categories were made out of concern for differences between cases within the group containing the largest facilities.

Prior to each hot-deck imputation step, correlations and stepwise regressions were processed in order to determine the ordering of soft boundaries. The variables, type of ownership and categorized number of clients, were almost always included as soft boundary variables.

5.4.7 Regression Procedure for Multi-Modality Providers

The above procedures do not control for correlation across items by type of treatment for multi-modality facilities (TYPECARE = 5 or 6). Using treatment-level regression models for multi-treatment facilities attempted to control the relationships between items at the treatment level. This method was applied to admissions, revenues and cost items. The hot-deck was used for clients for single and multi-modality facilities since the missingness in the client's block was at modality levels not asked for the other 'blocks' (i.e., admissions, revenues, and costs). The staffing items were not broken out by type of care.

The regression imputation procedure was similar to what is discussed in Section 5.4.5, however the dependent variables and continuous auxiliary independent variables were proportions, in order to predict the within-facility allocation of the grand total (admissions, costs, and revenues), given the within-facility proportions of clients. No random error was added in order to control the predicted proportions between values of 0 and 1. A sequential method, similar to that used for the grand totals, was used to first fill-in admissions, then revenues, and then costs.

5.4.8 Evaluation of Imputed Data

After each regression and hot-deck step, edits, logical imputations, frequencies, and other summary statistics were run on the imputed values and also on the observed and imputed values so that the effect of imputation on univariate and joint distributions could be studied. The edits were processed after each imputation step to check the imputed values. The logical imputations were processed after each imputation step, in order to fill-in any missing values to which the imputed value could be logically deduced. The review of other summary statistics lead to revisions to the imputation process, for instance:

- An additional level of the categorized version of total admissions was created after the range of imputed number of SSI or SSDI clients admitted was considered too large;
- The amount of error added to predicted values for the total cost model (specifically, the model for hospital inpatient only) was reduced after univariate statistics showed that too much error was being added; and
- Regression imputation was implemented for items relating to types of care after initial procedures that included using hot-deck imputation did not maintain the correlations between clients, admissions, revenues, and costs at the type of care level.

The model variables, adjusted R^2 values for each model, and the hot-deck cells are presented in Section 5.5 for each imputation step.

5.5 Imputation Results

This section provides results of each statistical imputation by block of items. For each block of items, the list of boundary variables, or model variables, and adjusted R^2 values are presented. Variable descriptions for the imputation process variables are given in Appendix D.

5.5.1 Clients Block

The clients' block contained the trigger item D1 for item D4. The item D1 was not an imputation item; therefore, any missing values for D4 remained if D1 was missing. Table 5-6 shows the hard and soft boundaries for the client's block.

Table 5-6. Client block imputation

Item(s) to impute	Hard boundaries*	Soft boundaries	Basis of imputation
Client block	TYPECARE, B_HBOUND, B_BC, B_MM	B_D1**, BUCLNT2, OWN	Proportions/ratios from donor

*New variables were created as boundary variables in order to cut down on the number of hard boundary cells.

**The trigger item B-D1 was included as a soft boundary due to the low number of donors available within each hard boundary cells. The software would have failed because there were no available donors since hard boundaries could not be crossed for one record with a missing value.

5.5.2 Admissions

5.5.2.1 Total Admissions

Random regression was used to fill-in missing data for total admissions. This task followed the imputation for the client's block of items and preceded the use of the hot-deck for the admissions block of items. If sub-items were reported and the grand total was missing, then the regression imputation methods were not used. Instead, a donor was selected and its ratio of total admissions to its subtotal was applied to the donee's reported subtotal to arrive at the imputed total admissions.

Since the residuals from resulting models did not adhere to constant variance and normality, which is essential for adding error from a normal distribution, continuous variables were transformed using the natural log. Stepwise regressions were done in order to include only significant independent variables. Extreme outliers were removed if DFFITS[5] > 2 or the absolute value of the standardized residual was greater than 3.5. The resulting residual plots improved across types of care in general. Table 5-7 lists the significant independent variables for the admissions block, and the adjusted R^2 values.

[5] DFFITS is a measure of influence that a case has on the fitted value. The numerator is computed as the difference between the *i*-th case's fitted value using all cases in the model and its predicted value when removing cases *i* from the modeled data. The denominator involves a function of the mean square error when case *i* is omitted, so that the value of DFFITS for case *i* roughly represents the number of estimated standard deviations that the fitted value changes when case *i* is removed from model building.

Table 5-7. Results of the C2F1 imputation

Type of care	Independent variables	Adjusted R^2	Number of records imputed
1	LOGB1, CREG1, PSUT1, CREG3, PSUT2, CREG2	0.52	1
2	LOGB1, OWN2, CREG2, CREG3	0.33	0
3	LOGB1, CREG1, CREG2, CREG3, OWN1	0.43	1
4	LOGB1, PSUT2, CREG1	0.62	5
5	LOGB1, OWN1	0.40	2
6	LOGB1, OWN2, PSUT2, CREG2	0.61	0
Overall			9

5.5.2.2 Single Modality Providers

The number of admissions for the type of care offered by single modality providers was set equal to the total admissions.

5.5.2.3 Multi-Modality Providers

The allocation of the total admissions to the types of care provided for multi-modality facilities was imputed through regression models, using observed proportions of the number of clients, and/or other significant auxiliary variables. Results are given in the following Tables 5-8 and 5-9.

Table 5-8. Models for inpatient, residential, and outpatient proportions of total admissions

Model	Universe			Dependent variable	Independent variables	Adjusted R^2	Number of records imputed
	RB1A1	RB1D1	RB1G1				
1	1	1	1	PC2A1_2	PB1A2_2, CREG1	.43	1
2	1	1	1	PC2B1_2	PB1D2_2, CREG3	.71	1
3	1	2	1	PC2A1	PB1A2, PSUT1, CREG2	.51	2
4	2	1	1	PC2B1	PB1D2, CREG2, OWN1	.56	2

Table 5-9. Models for inpatient, residential, and outpatient proportions of outpatient admissions

Model	Universe		Dependent variable	Independent variables	Adjusted R^2	Number of records imputed
	RB1H1	RB1I1				
1	1	1	PC2D1	PB1H2, PSUT1	.85	2

5.5.2.4 Other Items (Not Associated With Type of Care)

The imputation of the remainder of the admissions block, required that we first impute for the trigger item C4A, then the trigger item C4B, so we can obtain complete data for C4ANUM and C4BNUM and use them as hard boundaries to impute for the rest of the admissions block. Table 5-10 shows the imputation steps and the hard and soft boundaries for each step within the admissions block. Each step required a selection of a donor.

Table 5-10. Hot-deck procedures for the admissions block

Item(s) to impute	Hard boundaries	Soft boundaries	Basis of imputation
C4A	TYPECARE, C_FEM	BUCLNT2, OWN, CENREG	Value from donor
C4B	TYPECARE	BUCLNT2, OWN, CENREG	Value from donor
C2 Column 1	TYPECARE	BUCLNT2, OWN	Donor's proportions/ratios
C4ANUM, C4BNUM	TYPECARE, C4A, C4B	C_FEM, CTC2F1, BUCLNT2	Donor's proportions/ratios

For the items C4A and C4ANUM, which ask for the number of pregnant females admitted, care was taken to ensure that the facility was not a facility for men only. As a further precaution, the facility needed to have reported female clients in order to have pregnant females admitted.

5.5.3 Staffing Block

5.5.3.1 Total Full-Time Staff

The continuous variables were transformed to impute for total full-time staff. This task follows the imputation for the admissions block of items and precedes the use of the hot-deck for the staffing block.

For the staffing block, respondents could report the number of staff in full-time equivalent units. This occurred 110 times. For these cases, the total full-time staff was be imputed so that the full vector of data for total full time staff could be used as a predictor variable for revenues and cost grand totals. Table 5-11 gives the list of independent variables for each type of care.

Table 5-11. Results of the A9I1 imputation

Type of care	Independent variables	Adjust R^2	Number of imputed values*
1	LOGB1 LOGC2 PSUT1	.48	16
2	LOGB1 LOGC2 OWN1 OWN2 PSUT1 CREG1 CREG2 CREG3	.50	14
3	LOGB1 LOGC2 OWN1 CREG1 CREG2	.68	4
4	LOGB1 LOGC2 OWN1 OWN2 PSUT2 CREG1 CREG2 CREG3	.39	56
5	LOGB1 LOGC2	.35	22
6	LOGB1 OWN1	.59	2
Overall			114

*Number of imputed values includes facilities reporting full time equivalents. These records are imputed for model building purposes only. Of the number of imputed values listed, two facilities in type of care category ‘2’, and two facilities in type of care category ‘4’ were imputed for missing staffing items.

5.5.3.2 Other Staffing Items

For imputing the remainder of the staffing block, Table 5-12 shows the hard and soft boundaries for the staffing block. Since all total full time staff were imputed by this stage, the categorized version of full time staff was used as a soft boundary.

Table 5-12. Hot-deck procedures for staffing items

Item(s) to impute	Hard boundaries	Soft boundaries	Basis of imputation
Remainder of staffing block	TYPECARE	CTA9I1, BUCLNT2, OWN	Donor's proportions/ratios

5.5.4 Revenue Block

5.5.4.1 Total Revenues

The procedure for building regression models for imputing for total revenues was similar to that of total admissions and total full time staff. Table 5-13 gives the list of independent variables for each type of care.

Table 5-13. Results of the D7 imputation

Type of care	Variables used	Adjusted R^2	Number of records imputed
1	LOGB1 LOGC2 CREG2	.69	29
2	LOGA9 LOGB1 LOGC2 OWN2 CREG1 CREG2	.78	10
3	LOGA9 LOGB1 LOGC2 OWN1 CREG1 CREG3	.79	21
4	LOGA9 LOGB1 LOGC2 OWN1 OWN2 PSUT1 PSUT2 CREG1 CREG2 CREG3	.65	68
5	LOGA9 LOGC2 OWN1 OWN2 PSUT1 PSUT2 CREG1	.57	24
6	LOGA9 LOGB1 LOGC2 OWN1	.72	1
Overall			153

5.5.4.2 Single Modality Providers

The amount of revenues for the type of care offered by single modality providers was set equal to the total revenues.

5.5.4.3 Multi-Modality Providers

The allocation of the total revenues to the types of care provided for multi-modality facilities were predicted through regression models, using observed proportions of the number of clients, and/or other significant auxiliary variables. The Tables 5-14 and 5-15 show the results.

Table 5-14. Models for inpatient, residential, and outpatient proportions of total revenues

	Universe			Dependent	Independent	Adjusted	Number of
Model	RB1A1	RB1D1	RB1G1	variables	variables	R^2	records imputed
1	1	1	1	PD12A	PB1A2, PC2B1, CREG1	.86	5
1	1	1	1	PD12B	PB1A2, OWN1, CREG3	.91	5
2	1	1	1	PD12A_3	PB1A2_3 PSUT1, PSUT2, CREG3, CREG1	.97	1
3	1	2	1	PD12A	PB1A2, PC2A1, CREG1, PSUT2	.36	13
4	2	1	1	PD12B	PB1D2, PC2B1	.52	5
Overall							29

Table 5-15. Models for inpatient, residential, and outpatient proportions of outpatient revenues

	Universe		Dependent	Independent	Adjusted	Number of
Model	RB1H1	RB1I1	variables	variables	R^2	records imputed
1	1	1	PD13A	PB1H2, PC2D1, CREG1	.91	11

5.5.5 Cost Block

5.5.5.1 Total Costs

The procedure for building regression models to impute for total costs was similar to that of total revenues, total admissions and total full time staff. Table 5-16 gives the list of independent variables for each type of care.

Table 5-16. Results of the D14 imputation

Type of care	Independent variables	Adjusted R^2	Number of records imputed
1	LOGREV LOGC2 CREG2 LOGA9 OWN1	.96	43
2	LOGREV LOGA9 LOGB1 OWN2	.99	17
3	LOGREV CREG1 LOGB1 OWN1 PSUT2 OWN2	.98	34
4	LOGREV OWN1 LOGA9 CREG3 LOGB1	.98	111
5	LOGREV LOGB1 OWN1 LOGA9	.94	39
6	LOGREV LOGA9 OWN2	.99	3
Overall			247

5.5.5.2 Single Modality Providers

The amount of costs for the type of care offered by single care providers were imputed through the use of random-within-cells hot-deck using the proportional allocation of the donor applied to the donee's total costs. Because of the extra sub-item that asks for costs not attributable by type of care, we cannot directly set the type of care cost equal to the total costs. Resulting hot-deck cells are shown in Table 5-17.

Table 5-17. Results of the hot-deck procedures for the cost block

Item(s) to impute	Hard boundaries	Soft boundaries	Basis of imputation
D14, D15 items, D16D	TYPECARE	CTD7, BUCLNT2, OWN	Donor's proportions/ratios

5.5.5.3 Multi-Modality Providers

The allocation of the total costs to the types of care provided for multi-modality providers was imputed through regression models, using observed proportions of the number of clients, and/or other significant auxiliary variables. Summary Tables 5-18 and 5-19 are provided below.

Table 5-18. Models for inpatient, residential, and outpatient proportions of total costs

	Universe			Dependent	Independent	Adjusted	Number of
Model	RB1A1	RB1D1	RB1G1	variable	variables	R^2	records imputed
1	1	1	1	PD16A	PD12A, PB1D2, OWN2	.94	6
1	1	1	1	PD16B	PD12B, PSUT1	.95	6
2	1	2	1	PD16A	PB1A2, PC2A1, PD12A	.90	33
3	2	1	1	PD16B	PB1D2, PD12B	.91	7
Overall							52

Table 5-19. Models for inpatient, residential, and outpatient proportions of outpatient costs

	Universe		Dependent	Independent	Adjusted	Number of
Model	RB1H1	RB1I1	variables	variables	R^2	records imputed
1	1	1	PD16C2	PB1H2, PD13A	.95	14

5.5.5.4 Other Items (Not Associated With Type of Care)

The other items in the cost block are the break out of total costs to 3 expense categories. The same donor that was selected for the single care providers was used to fill in the three expense categories.

5.6 Some Ratio Outliers

There are a small number of imputed values that have resulted as being outliers in terms of certain ratios. There are thirteen records that have been flagged using the variable RATIOFLG. The thirteen records each have ratio outliers in at least one of the following ratios: D16C2/B1H2, C2A1/B1A2, D13B/C2E1, D13B/B1I2, D16C1/B1I2, D12A/C2A1, D16A/C2A1, D16B/C2B1, D12B/D16B, D14/B1J2, D14/A9I1, D7/A9I1, D7/B1J2, D16A/B1A2.

The imputation methods were constructed to generate imputed values for the full range of values that were observed in the reported data. Therefore, one would expect to have ratio 'outliers' in terms of the top 10 (and bottom 10) ratio values for each ratio, such as those mentioned in the previous paragraph. One goal of the imputation procedure was to have the imputed values reflect what was reported in the data.

5.7 Imputation Flag Variables

The imputation file contains imputation flags, which identify the source of the imputed value. For each imputation variable, there is an imputation flag variable, which is named as '*original variable name*_F', except for the variables B2INE10, B2REE10, B2OME10, B2ONE10, which were renamed to B2INE10F, B2REE10F, B2OME10F, B2ONE10F, respectively. Values of the imputation flag variables are:

0 = No action taken for the item.
3 = Missing value filled-in using logical imputations from a simple difference function (the difference between the total and the non-missing sub-items, and will be applied to where there is one missing sub-item).
4 = Missing value filled-in using 1996 UFDS.
5 = Missing value filled-in using Phase II data.
6 = Missing value filled-in using regression imputation for multi-modality facilities.
7 = Missing value filled-in using random regression imputation for grand totals for blocks, for single modality facilities and for multi-modality facilities with one missing sub-item.
8 = Missing value filled-in using hot-deck imputation.
9 = Missing value filled-in using 1997 UFDS.
10 = Copied data from another item.
11 = Dollar values converted from a reported percentage.
12 = Dollar values converted from a reported percentage, then rounded to make the sum of parts add to the total. This code was also applied to the items in B2, where imputed values were rounded to make the sum of parts add to the total.
13 = Percentages converted from imputed or observed dollar values.
14 = Percentages converted from a dollar value, then rounded to make the sum of the parts add to 100.

When logical imputation was used during a process that assigned an imputation flag = 6, 7, or 8, then the imputation flag = 6, 7, or 8. For example, suppose total revenues was imputed using random regression, then logical imputation was used to fill-in the hospital inpatient revenues, since it was the only treatment offered, then the imputation flag associated with the hospital inpatient revenues item was coded to reflect the random regression imputation. However, if data values were copied or reported percentages were used to fill-in data, then the imputation flag = 10, 11, or 12 to reflect the way the imputed values were transferred or converted. To find the true source of an item value associated with the flag values of 10, 11, or 12, one needs to find the flag value for the associated grand total. For instance, suppose one may notice a flag value of 11 for hospital inpatient revenues, which means that the reported percentage of total revenues was used to impute the dollar value. One can find out if the total revenue value was imputed for that record by checking its imputation flag.

5.8 Converting to Substance Abuse Only

When D10BOX = 1, the respondent reported revenues beyond revenues for their substance abuse treatment program, and D10PC is the percentage attributable to substance abuse. However, there are times that D10BOX = 1 but total revenues (D7) was missing (e.g., the respondent reported percentages in D12 and D13 but no total). We imputed total revenues for substance abuse only; therefore, whenever D7 is imputed (i.e., FD7 = 4, 5, 7, 8, or 9), then D10BOX and D10PC should be ignored. The same is true for D14 (total costs), and its corresponding variables D18BOX and D18PC.

5.9 Remaining Missing Values

Missing values still remain for some imputation items. The item D4 (percent covered by managed care) remained missing when item D1 (trigger) was missing. Several missing values remain for the source of revenues questions (D8 items) since statistical imputation methods were not applied.

5.10 Impact of Imputation

If item nonresponse occurred at random then no distributional differences between the observed values and the full data vector, which includes observed and imputed values, are expected. Unweighted means and standard deviations were computed for items with relatively high amounts of missingness. The results are given in Table 5-20. In addition, weighted means show the impact when imputed values are weighted. It should be noted that the values for the financial items were transformed in order to represent costs and revenues attributable to substance abuse only.

The table shows small differences between the unweighted means for observed values only and all records, ranging from –3 percent to 4 percent for most items. The largest difference, 10 percent, occurs for item C4BNUM (number of pregnant females). Further investigation showed that the missingness did not occur at random since total admissions was approximately 7 percent higher among all records, when compared to observed records only. Differences between unweighted means for other items could just as well be explained by the imputation models reflecting the fact that one cannot ignore the missing data mechanism.

Differences between the unweighted standard deviations can be a function of several reasons. One of which is the removal of several records from the model fitting process. Another reason could be due to the extent that the residuals distribution followed a normal distribution. A further reason is that imputed values that contribute to estimates for a particular domain, most likely were generated from more than one model, since models were generated by the variable TYPECARE.

Table 5-20. Unweighted means, standard deviations and weighted means of selected items*

				Unweighted means			Unweighted standard deviations			Weighted means		
Item	Total observed	Total imputed and observed	Remaining missing values	Observed	All records	Percent difference (all - obs)	Observed	All records	Percent difference (all - obs)	Observed	All records	Percent difference (all - obs)
Items pertaining to all facilities												
C2F1	2374	2395		446	446	0%	754	752	0%	346	346	0%
C4A	2344	2395		1	1	0%	0	0	0%	1	1	2%
C4ANUM	2245	2395		5	5	4%	15	15	0%	4	4	5%
C4B	2297	2395		1	1	0%	0	0	0%	1	1	2%
C4BNUM	2101	2395		39	43	10%	103	107	4%	27	32	14%
D7	2169	2395		1,003,574	1,004,449	0%	2,187,910	2,151,368	-2%	667,925	675,068	1%
D14	2141	2395		872,964	882,745	1%	1,388,190	1,424,314	3%	587,165	604,849	3%
D15A	2124	2395		560,130	570,751	2%	941,032	979,107	4%	376,171	388,458	3%
D15B	2126	2395		33,330	33,976	2%	110,053	109,133	-1%	25,049	28,259	11%
D15C	2135	2395		274,631	277,132	1%	534,356	534,924	0%	181,985	187,304	3%
D16D	2233	2395		20,235	20,660	2%	224,179	224,913	0%	18,694	18,479	-1%
Items pertaining to inpatient facilities												
D12A	271	343		1,987,739	1,994,909	0%	2,349,371	2,402,090	2%	1,391,579	1,417,465	2%
D16A	258	343		1,530,538	1,544,533	1%	2,142,454	2,200,300	3%	1,038,084	1,129,068	8%
Items pertaining to residential facilities												
D12B	565	598		1,257,212	1,251,931	0%	3,451,278	3,365,303	-3%	874,111	898,527	3%
D16B	567	598		1,038,286	1,048,570	1%	1,424,235	1,412,007	-1%	758,451	792,055	4%
Items pertaining to all outpatient facilities												
D12C	1586	1761		555,719	539,983	-3%	823,055	797,944	-3%	377,140	373,670	-1%
D16C	1560	1761		508,926	495,298	-3%	837,709	809,153	-4%	330,524	328,633	-1%
Items pertaining to outpatient methadone facilities												
D13A	376	418		874,970	854,342	-2%	792,606	780,355	-2%	776,382	773,277	0%
D16C2	369	418		860,009	838,412	-3%	780,032	769,510	-1%	751,615	752,654	0%
Items pertaining to residential facilities												
D13B	1285	1435		418,922	409,725	-2%	765,022	738,544	-4%	333,626	333,460	0%
D16C1	1262	1435		383,462	374,857	-2%	787,163	754,907	-4%	293,250	293,948	0%

*All statistics relating to financial items were computed based on data that were converted to represent substance abuse treatment only (see Section 5.12).

Differences between weighted and unweighted means have two major causes. It shows the effect due to probability proportionate to size sampling, that is, larger facilities had a higher chance of selection, which gives them a lower weight. This explains why the weighted means are about one-third less than the unweighted means. Observing the differences between the observed weighted means and the weighted means using all records shows another reason. The differences show the impact of applying the weights to the imputed values.

5.11 Measuring Imputation Error Variance

Treating imputed values as if they had actually been observed or reported may lead to a significant understatement of the variance of the estimate. Several methods have been developed to account for the effects of imputation error in variance estimation. Such methods include, but are not limited to, applying an adjustment to jackknife replicates (Rao and Shao (1992)); a model-assisted approach (Lee, Rancourt, and Sarndal (1995)); a bootstrap method (Shao and Sitter (1996)), and using multiple imputations (Rubin (1977) and recently by Schafer (1997)). The All-Cases Imputation (ACI) method was developed recently (Montaquila and Jernigan (1997)) and we chose to apply the method to ADSS (Krenzke, Mohadjer, Montaquila, 1998).

Since the missingness rates in the clients, staffing, and much of the admissions blocks was small, we assume that the imputation error variance is negligible for these items. For selected items with higher nonresponse rates (total revenues, total costs, and other key items within the admissions, revenue, and cost blocks), the imputation error variance was estimated using the ACI method. The ACI method involves imputing for all cases, not just nonrespondents, and using the imputed and observed values for the respondents to estimate the imputation error among the nonrespondents. The model-assisted ACI approach assumes ignorable nonresponse and it provides an unbiased estimate of the variance of the mean under generalized conditions. In general, the ACI estimator of the total variance has three components. The first component is the sampling error variance (S^2), the second component is the imputation error variance (I^2), and the third component is the imputation error covariance. The third component is considered negligible for ADSS since the donors were allowed to be used only once. Therefore, the ACI estimator of the variance of the mean reduces to two terms:

$$\hat{v}_{ACI}\left(\bar{y}_{st}\right) = \sum_{h=1}^{L}\left(\frac{N_h}{N}\right)^2\left[\frac{(1-f_h)}{n_h}\hat{v}\left(y_{hi}^{*}\right) + \frac{m_h}{n_h^2}\hat{v}\left(\tau_{hi}\right)\right];$$

where, $\hat{v}(y^*_{hi})$ = the sample variance among the actual and imputed values of the characteristic y in stratum h, and $\hat{v}(\tau_{hi})$ = the sample variance among the respondent imputation errors in stratum h. Both variance terms were computed using WesVar Complex Samples 3.0, a software package for analyzing complex samples. The stratified jackknife technique, using the replication scheme discussed in Section 4, was used to compute the variance components.

To simplify the computation of variances in the presence of imputation error, the approach recommended for ADSS was to incorporate the imputation error variance by using a variance inflation factor (VIF). The variance inflation factor can be multiplied by the variance (computed by treating imputed values as if they were observed) after the calculation of the jackknife variances. This procedure is done in the same manner as a design effect being applied to a variance whose computation used a formula for simple random sampling.

Table 5-21 shows the results for each of the items that the imputation error variance was directly computed. Other VIFs, as noted, were generalized through simply using the ratio of the VIF to the imputation rate, for a closely related item. The amount of missingness filled-in by statistical imputation methods is shown, as well as the variance inflation factor (VIF), which is VIF = $(S^2 + I^2)/S^2$. For financial items, variances were computed using values converted to represent substance abuse only.

Table 5-21. VIFs and statistical imputation rates

Item	S*	I	VIF	Statistical imputation rate (%)
Items not relating to type of care				
C4A			1.06**	2.13
C4ANUM			1.04	6.26
C4B	0.02	0.01	1.11	4.09
C4BNUM	3.23	0.85	1.07	12.28
D7	26,958.35	3,635.05	1.02	6.43
D14	24,287.61	2,165.58	1.01	10.35
D15A	15,712.60	2,546.83	1.03	11.32
D15B			1.03***	11.23
D15C			1.03	11.23
D16D	5114.76	1,451.02	1.08	6.76
Items relating to hospital inpatient treatment				
D12A	83,707.99	22,065.09	1.07	17.20
D16A	74,972.70	37,205.35	1.25	24.49
Items relating to residential treatment				
D12B	61,260.51	5,725.108	1.01	4.35
D16B	43,904.80	3,834.453	1.01	5.02
Items relating to outpatient treatment				
D12C	18,267.48	3,179.23	1.03	7.10
D16C	18,234.17	1,097.14	1.00	11.24
Items not relating to type of care				
D13A			1.03	7.65
D16C2			1.00	11.48
Items relating to methadone treatment				
D13B	19,433.75	3,444.70	1.03	7.66
D16C1	19,244.03	1,125.07	1.00	11.84

*S and I were computed after converting the financial items so that they represent substance abuse treatment only (see note in Section 5.12).

**The VIFs were not directly computed for C4A and C4ANUM. The ratio of the VIF to statistical imputation rate for the corresponding C4B terms was used to approximate the VIF for the C4A items.

***The VIFs were not directly computed for D15B and D15B. The ratio of the VIF to statistical imputation rate for D15A was used to approximate the VIF for D15B and D15C.

5.12 Note to Data Analysis

When D10BOX = 1, the respondent reported revenues beyond revenues for their substance abuse treatment program, and D10PC is the percentage attributable to substance abuse. However, there are times that D10BOX = 1 but total revenues was missing (e.g., the respondent reported percentages in D12 and D13 but no total). We imputed total revenues for substance abuse only, therefore, whenever D7 is imputed (i.e., FD7 = 4, 5, 7, 8, or 9), then D10BOX and D10PC should be ignored. The same is true for D14 (total costs), and its corresponding variables D18BOX and D18PC.

REFERENCES

Cochran, W.G. (1977). Sampling techniques, 3rd edition, New York: John Wiley.

Judkins, D. (1997). Imputing for swiss cheese patterns of missing data. To appear in the *Statistics Canada Symposium 97 Proceedings*.

Kalton, G. and Kish, L. (1984). Some efficient random imputation methods. *Communications in Statistics*, 13(16), pp. 1919-1939.

Krenzke, T., Mohadjer, L., and Montaquila, J. (1998). Generalizing the imputation error variance in the alcohol and drug services study. To appear in the *American Statistical Association's 1998 Proceedings of the Section on Biometrics*.

Lee, H., Rancourt, E., and Sarndal, C.E. (1995). Variance estimation in the presence of imputed data for the generalized imputation system. *American Statistical Association's 1995 Proceedings of the Section on Survey Research Methods,* pp. 384-389.

Montaquila, J. and Jernigan, R. (1997). Variance estimation in the presence of imputed data. *American Statistical Association's 1997 Proceedings of the Section on Survey Research Methods,* pp. 273-278.

Montaquila, J. and Ponikowski, C. (1995). An evaluation of alternative imputation methods. *American Statistical Association's 1995 Proceedings of the Section on Survey Research Methods,* pp. 281-286.

Ohlsoon, E. (1995). Coordination of samples using permanent random numbers. Business Survey Methods. Cox, B.G. et al Editors, pp. 153-169.

Rao, J.N.K. and Shao, J. (1992). Jackknife variance estimation with survey data under hot-deck imputation. *Biometrika*, 79, pp. 811-822.

Rubin, D.B. (1987). Multiple imputation for nonresponse in surveys, New York: John Wiley.

Schafer, J. (1997). Analysis of incomplete multivariate data. Chapman and Hall

Shao, J. and Sitter, R.R. (1996). *Bootstrap for Imputed Survey Data.* Technical Report 227, Carleton University, Laboratory for Research in Statistics and Probability.

Wolter, K.M. (1985). Introduction to variance estimation. New York: Springer-Verlag.

APPENDIX A

(Not Available)

APPENDIX B

ITEM DESCRIPTIONS

Item Descriptions

Item	Item description
Items pertaining to all facilties	
A9A1	Number of full-time doctors on staff.
A9B1	Number of full-time registered nurses on staff.
A9C1	Number of full-time other medical personnel on staff (LPN, PA, etc.).
A9D1	Number of full-time doctor level counselors on staff (Psychologists, etc.)
A9E1	Number of full-time masters level counselors (MSW, etc.)
A9F1	Number of full-time other degreed counselors on staff.
A9G1	Number of non-degreed counselors on staff.
A9H1	All other full-time staff including administrative staff.
A9I1	Total number of full-time staff.
A9A2	Number of part-time doctors on staff.
A9B2	Number of part-time registered nurses on staff.
A9C2	Number of part-time other medical personnel on staff (LPN, PA, etc.).
A9D2	Number of part-time doctor level counselors on staff (Psychologists, etc.)
A9E2	Number of part-time masters level counselors (MSW, etc.)
A9F2	Number of part-time other degreed counselors on staff.
A9G2	Number of non-degreed counselors on staff.
A9H2	All other part-time staff including administrative staff.
A9I2	Total number of part-time staff.
A9A3	Number of contract and consultant doctors on staff.
A9B3	Number of contract and consultant registered nurses on staff.
A9C3	Number of contract and consultant other medical personnel on staff (LPN, PA, etc.).
A9D3	Number of contract and consultant doctor level counselors on staff (Psychologists, etc.)
A9E3	Number of contract and consultant masters level counselors (MSW, etc.)
A9F3	Number of contract and consultant other degreed counselors on staff.
A9G3	Number of non-degreed counselors on staff.
A9H3	All other contract and consultant staff including administrative staff.
A9I3	Total number of contract and consultant staff.
B1J2	Total active substance abuse clients on October 1, 1996.
B1J3	Total active substance abuse clients on October 1, 1996 who were dispensed methadone.
B3	Number of outpatient substance abuse clients visiting the facility for treatment during the week ending October 4, 1996.
B4	Number of outpatient visits for substance abuse services during the week ending October 4, 1996.
B12A	Number of clients in detoxification using methadone.
B12B	Number of clients in methadone maintenance.
C2F1	Total number of admissions during the most recent 12-month period.
C4A	Did the facility admit any pregnant females during the most recent 12-month period for substance abuse services?
C4ANUM	The number of pregnant females admitted during the most recent 12-month period for substance abuse services.
C4B	Did the facility admit any SSI or SSDI clients during the most recent 12-month period for substance abuse services?
C4BNUM	The number of SSI or SSDI clients admitted during the most recent 12-month period for substance abuse services.
D1	Did the facility have any formal written agreements with any managed care organizations for substance abuse treatment?
D4	The number of managed care organizations did the facility have contract arrangements for substance abuse treatment.
D7	Total substance abuse treatment revenue or funding.
D8A	Amount of substance abuse revenue from client fees.
D8B	Amount of substance abuse revenue from private health insurance, fee-for-service.
D8C	Amount of substance abuse revenue from private health insurance, HMO/PPO/Managed Care.
D8D	Amount of substance abuse revenue from Medicaid, not specified.
D8E	Amount of substance abuse revenue from Medicaid managed care.
D8F	Amount of substance abuse revenue from Medicare.
D8G	Amount of substance abuse revenue from other federal government funds (VA, CHAMPUS, etc.).
D8H	Amount of substance abuse revenue from other public funds.
D8I	Amount of substance abuse revenue from other sources.
D8APC	Percentage of substance abuse revenue from client fees.
D8BPC	Percentage of substance abuse revenue from private health insurance, fee-for-service.
D8CPC	Percentage of substance abuse revenue from private health insurance, HMO/PPO/Managed Care.
D8DPC	Percentage of substance abuse revenue from Medicaid, not specified.
D8EPC	Percentage of substance abuse revenue from Medicaid managed care.
D8FPC	Percentage of substance abuse revenue from Medicare.
D8GPC	Percentage of substance abuse revenue from other federal government funds (VA, CHAMPUS, etc.).
D8HPC	Percentage of substance abuse revenue from other public funds.
D8IPC	Percentage of substance abuse revenue from other sources.
D14	Total substance abuse treatment costs.
D15A	Amount of substance abuse treatment cost for employee personnel expenses.

Item Descriptions

Item	Item description

Items pertaining to all facilties (continued)

D15B	Amount of substance abuse treatment cost for other personnel expenses.
D15C	Amount of substance abuse treatment cost for nonpersonnel expenses.
D15APC	Percentage of substance abuse treatment cost for employee personnel expenses.
D15BPC	Percentage of substance abuse treatment cost for other personnel expenses.
D15CPC	Percentage of substance abuse treatment cost for nonpersonnel expenses.
D16D	Amount of substance abuse treatment costs not attributable by type of care

Items pertaining to inpatient facilties

B2INA1	Number of male inpatient clients as of October 1, 1996.
B2INA2	Number of female inpatient clients as of October 1, 1996.
B2INA3	Number of inpatient clients whose sex is unknown as of October 1, 1996.
B2INB1	Number of White (not Hispanic) inpatient clients as of October 1, 1996.
B2INB2	Number of Black (not Hispanic) inpatient clients as of October 1, 1996.
B2INB3	Number of Hispanic inpatient clients as of October 1, 1996.
B2INB4	Number of Asian or Pacific Islander inpatient clients as of October 1, 1996.
B2INB5	Number of American Indian or Alaskan Native inpatient clients as of October 1, 1996.
B2INB6	Number of inpatient clients of unknown race/ethnicity as of October 1, 1996.
B2INC1	Number of inpatient clients under 18 years old at admission as of October 1, 1996.
B2INC2	Number of inpatient clients 18-24 years old at admission as of October 1, 1996.
B2INC3	Number of inpatient clients 25-34 years old at admission as of October 1, 1996.
B2INC4	Number of inpatient clients 35-44 years old at admission as of October 1, 1996.
B2INC5	Number of inpatient clients 45 years or older at admission as of October 1, 1996.
B2INC6	Number of inpatient clients with unknown age at admission as of October 1, 1996.
B2IND1	Number of inpatient clients whose primary source of payment was no payment as of October 1, 1996.
B2IND2	Number of inpatient clients whose primary source of payment was self payment as of October 1, 1996.
B2IND3	Number of inpatient clients whose primary source of payment was fee-for-service private health insurance as of October 1, 1996.
B2IND4	Number of inpatient clients whose primary source of payment was HMO/PPO/managed care private health insurance as of October 1, 1996.
B2IND5	Number of inpatient clients whose primary source of payment was Medicaid as of October 1, 1996.
B2IND6	Number of inpatient clients whose primary source of payment was Medicare as of October 1, 1996.
B2IND7	Number of inpatient clients whose primary source of payment was other public payment as of October 1, 1996.
B2IND8	Number of inpatient clients whose primary source of payment was unknown as of October 1, 1996.
B2INE1	Number of inpatient clients whose principal drug of abuse was heroin/other opiates as of October 1, 1996.
B2INE2	Number of inpatient clients whose principal drug of abuse was cocaine as of October 1, 1996.
B2INE3	Number of inpatient clients whose principal drug of abuse was benzodiazepines as of October 1, 1996.
B2INE4	Number of inpatient clients whose principal drug of abuse was barbiturates as of October 1, 1996.
B2INE5	Number of inpatient clients whose principal drug of abuse was amphetamines as of October 1, 1996.
B2INE6	Number of inpatient clients whose principal drug of abuse was marijuana/hashish/THC as of October 1, 1996.
B2INE7	Number of inpatient clients whose principal drug of abuse was PCP/LSD as of October 1, 1996.
B2INE8	Number of inpatient clients whose principal drug of abuse was alcohol as of October 1, 1996.
B2INE9	Number of inpatient clients whose principal drug of abuse was other drugs (not alcohol) as of October 1, 1996.
B2INE10	Number of inpatient clients whose principal drug of abuse was unknown as of October 1, 1996.
B1A2	Total active substance abuse hospital inpatient clients on October 1, 1996.
B1B2	Total active substance abuse hospital inpatient detoxification clients on October 1, 1996.
B1C2	Total active substance abuse hospital inpatient rehabilitation clients on October 1, 1996.
B1A3	Total active substance abuse hospital inpatient clients on October 1, 1996 who were dispensed methadone.
B1B3	Total active substance abuse hospital inpatient detoxification clients on October 1, 1996 who were dispensed methadone.
B1C3	Total active substance abuse hospital inpatient rehabilitation clients on October 1, 1996 who were dispensed methadone.
C2A1	Number of total hospital inpatient admissions in the most recent 12-month period.
D12A	Amount of substance abuse revenue generated by hospital inpatient care.
D12APC	Percentage of substance abuse revenue generated by hospital inpatient care.
D16A	Amount of substance abuse costs incurred by hospital inpatient care.

Items pertaining to residential facilties

B1D2	Total active substance abuse residential clients on October 1, 1996.
B1E2	Total active substance abuse residential detoxification clients on October 1, 1996.
B1F2	Total active substance abuse residential rehabilitation clients on October 1, 1996.
B1D3	Total active substance abuse residential clients on October 1, 1996 who were dispensed methadone.
B1E3	Total active substance abuse residential detoxification clients on October 1, 1996 who were dispensed methadone.
B1F3	Total active substance abuse residential rehabilitation clients on October 1, 1996 who were dispensed methadone.
B2REA1	Number of male residential clients as of October 1, 1996.

Item Descriptions

Item	Item description
Items pertaining to residential facilties (continued)	
B2REA2	Number of female residential clients as of October 1, 1996.
B2REA3	Number of residential clients whose sex is unknown as of October 1, 1996.
B2REB1	Number of White (not Hispanic) residential clients as of October 1, 1996.
B2REB2	Number of Black (not Hispanic) residential clients as of October 1, 1996.
B2REB3	Number of Hispanic residential clients as of October 1, 1996.
B2REB4	Number of Asian or Pacific Islander residential clients as of October 1, 1996.
B2REB5	Number of American Indian or Alaskan Native residential clients as of October 1, 1996.
B2REB6	Number of residential clients of unknown race/ethnicity as of October 1, 1996.
B2REC1	Number of residential clients under 18 years old at admission as of October 1, 1996.
B2REC2	Number of residential clients 18-24 years old at admission as of October 1, 1996.
B2REC3	Number of residential clients 25-34 years old at admission as of October 1, 1996.
B2REC4	Number of residential clients 35-44 years old at admission as of October 1, 1996.
B2REC5	Number of residential clients 45 years or older at admission as of October 1, 1996.
B2REC6	Number of residential clients with unknown age at admission as of October 1, 1996.
B2RED1	Number of residential clients whose primary source of payment was no payment as of October 1, 1996.
B2RED2	Number of residential clients whose primary source of payment was self payment as of October 1, 1996.
B2RED3	Number of residential clients whose primary source of payment was fee-for-service private health insurance as of October 1, 1996.
B2RED4	Number of residential clients whose primary source of payment was HMO/PPO/managed care private health insurance as of October 1, 1996.
B2RED5	Number of residential clients whose primary source of payment was Medicaid as of October 1, 1996.
B2RED6	Number of residential clients whose primary source of payment was Medicare as of October 1, 1996.
B2RED7	Number of residential clients whose primary source of payment was other public payment as of October 1, 1996.
B2RED8	Number of residential clients whose primary source of payment was unknown as of October 1, 1996.
B2REE1	Number of residential clients whose principal drug of abuse was heroin/other opiates as of October 1, 1996.
B2REE2	Number of residential clients whose principal drug of abuse was cocaine as of October 1, 1996.
B2REE3	Number of residential clients whose principal drug of abuse was benzodiazepines as of October 1, 1996.
B2REE4	Number of residential clients whose principal drug of abuse was barbiturates as of October 1, 1996.
B2REE5	Number of residential clients whose principal drug of abuse was amphetamines as of October 1, 1996.
B2REE6	Number of residential clients whose principal drug of abuse was marijuana/hashish/THC as of October 1, 1996.
B2REE7	Number of residential clients whose principal drug of abuse was PCP/LSD as of October 1, 1996.
B2REE8	Number of residential clients whose principal drug of abuse was alcohol as of October 1, 1996.
B2REE9	Number of residential clients whose principal drug of abuse was other drugs (not alcohol) as of October 1, 1996.
B2REE10	Number of residential clients whose principal drug of abuse was unknown as of October 1, 1996.
C2B1	Number of residential admissions in the most recent 12-month period.
D12B	Amount of substance abuse revenue generated by residential care.
D12BPC	Percentage of substance abuse revenue generated by residential care.
D16B	Amount of substance abuse costs incurred by residential care.
Items pertaining to all outpatient facilties	
B1G2	Total active substance abuse outpatient clients on October 1, 1996.
B1G3	Total active substance abuse outpatient clients on October 1, 1996 who were dispensed methadone
C2C1	Number of total outpatient admissions in the most recent 12-month period.
C2D1	Number of total outpatient methadone admissions in the most recent 12-month period.
C2E1	Number of total outpatient non-methadone admissions in the most recent 12-month period.
D12C	Amount of substance abuse revenue generated by outpatient care.
D12CPC	Percentage of substance abuse revenue generated by outpatient care.
D16C	Amount of substance abuse costs incurred by outpatient care.
Items pertaining to outpatient methadone facilities	
B1H2	Total active substance abuse outpatient methadone clients on October 1, 1996.
B1H3	Total active substance abuse outpatient methadone clients on October 1, 1996 who were dispensed methadone.
B2OMA1	Number of male outpatient methadone clients as of October 1, 1996.
B2OMA2	Number of female outpatient methadone clients as of October 1, 1996.
B2OMA3	Number of outpatient methadone clients whose sex is unknown as of October 1, 1996.
B2OMB1	Number of White (not Hispanic) outpatient methadone clients as of October 1, 1996.
B2OMB2	Number of Black (not Hispanic) outpatient methadone clients as of October 1, 1996.
B2OMB3	Number of Hispanic outpatient methadone clients as of October 1, 1996.
B2OMB4	Number of Asian or Pacific Islander outpatient methadone clients as of October 1, 1996.
B2OMB5	Number of American Indian or Alaskan Native outpatient methadone clients as of October 1, 1996.
B2OMB6	Number of outpatient methadone clients of unknown race/ethnicity as of October 1, 1996.

Item	Item description
Items pertaining to outpatient methadone facilities (continued)	
B2OMC1	Number of outpatient methadone clients under 18 years old at admission as of October 1, 1996.
B2OMC2	Number of outpatient methadone clients 18-24 years old at admission as of October 1, 1996.
B2OMC3	Number of outpatient methadone clients 25-34 years old at admission as of October 1, 1996.
B2OMC4	Number of outpatient methadone clients 35-44 years old at admission as of October 1, 1996.
B2OMC5	Number of outpatient methadone clients 45 years or older at admission as of October 1, 1996.
B2OMC6	Number of outpatient methadone clients with unknown age at admission as of October 1, 1996.
B2OMD1	Number of outpatient methadone clients whose primary source of payment was no payment as of October 1, 1996.
B2OMD2	Number of outpatient methadone clients whose primary source of payment was self payment as of October 1, 1996.
B2OMD3	Number of outpatient methadone clients whose primary source of payment was fee-for-service private health insurance as of October 1, 1996.
B2OMD4	Number of outpatient methadone clients whose primary source of payment was HMO/PPO/managed care private health insurance as of October 1, 1996.
B2OMD5	Number of outpatient methadone clients whose primary source of payment was Medicaid as of October 1, 1996.
B2OMD6	Number of outpatient methadone clients whose primary source of payment was Medicare as of October 1, 1996.
B2OMD7	Number of outpatient methadone clients whose primary source of payment was other public payment as of October 1, 1996.
B2OMD8	Number of outpatient methadone clients whose primary source of payment was unknown as of October 1, 1996.
B2OME1	Number of outpatient methadone clients whose principal drug of abuse was heroin/other opiates as of October 1, 1996.
B2OME2	Number of outpatient methadone clients whose principal drug of abuse was cocaine as of October 1, 1996.
B2OME3	Number of outpatient methadone clients whose principal drug of abuse was benzodiazepines as of October 1, 1996.
B2OME4	Number of outpatient methadone clients whose principal drug of abuse was barbiturates as of October 1, 1996.
B2OME5	Number of outpatient methadone clients whose principal drug of abuse was amphetamines as of October 1, 1996.
B2OME6	Number of outpatient methadone clients whose principal drug of abuse was marijuana/hashish/THC as of October 1, 1996.
B2OME7	Number of outpatient methadone clients whose principal drug of abuse was PCP/LSD as of October 1, 1996.
B2OME8	Number of outpatient methadone clients whose principal drug of abuse was alcohol as of October 1, 1996.
B2OME9	Number of outpatient methadone clients whose principal drug of abuse was other drugs (not alcohol) as of October 1, 1996.
B2OME10	Number of outpatient methadone clients whose principal drug of abuse was unknown as of October 1, 1996.
C2D1	Number of total outpatient methadone admissions in the most recent 12-month period.
D13A	Amount of substance abuse revenue generated by outpatient methadone care.
D13APC	Percentage of substance abuse revenue generated by outpatient methadone care.
D16C2	Amount of substance abuse costs incurred by outpatient methadone care.
Items pertaining to outpatient non-methadone facilities	
B1I2	Total active substance abuse outpatient non-methadone clients on October 1, 1996.
B1I3	Total active substance abuse outpatient non-methadone clients on October 1, 1996 who were dispensed methadone.
B2ONA1	Number of male outpatient non-methadone clients as of October 1, 1996.
B2ONA2	Number of female outpatient non-methadone clients as of October 1, 1996.
B2ONA3	Number of outpatient non-methadone clients whose sex is unknown as of October 1, 1996.
B2ONB1	Number of White (not Hispanic) outpatient non-methadone clients as of October 1, 1996.
B2ONB2	Number of Black (not Hispanic) outpatient non-methadone clients as of October 1, 1996.
B2ONB3	Number of Hispanic outpatient non-methadone clients as of October 1, 1996.
B2ONB4	Number of Asian or Pacific Islander outpatient non-methadone clients as of October 1, 1996.
B2ONB5	Number of American Indian or Alaskan Native outpatient non-methadone clients as of October 1, 1996.
B2ONB6	Number of outpatient non-methadone clients of unknown race/ethnicity as of October 1, 1996.
B2ONC1	Number of outpatient non-methadone clients under 18 years old at admission as of October 1, 1996.
B2ONC2	Number of outpatient non-methadone clients 18-24 years old at admission as of October 1, 1996.
B2ONC3	Number of outpatient non-methadone clients 25-34 years old at admission as of October 1, 1996.
B2ONC4	Number of outpatient non-methadone clients 35-44 years old at admission as of October 1, 1996.
B2ONC5	Number of outpatient non-methadone clients 45 years or older at admission as of October 1, 1996.
B2ONC6	Number of outpatient non-methadone clients with unknown age at admission as of October 1, 1996.
B2OND1	Number of outpatient non-methadone clients whose primary source of payment was no payment as of October 1, 1996.
B2OND2	Number of outpatient non-methadone clients whose primary source of payment was self payment as of October 1, 1996.
B2OND3	Number of outpatient non-methadone clients whose primary source of payment was fee-for-service private health insurance as of October 1, 1996.
B2OND4	Number of outpatient non-methadone clients whose primary source of payment was HMO/PPO/managed care private health insurance as of October 1, 1996.
B2OND5	Number of outpatient non-methadone clients whose primary source of payment was Medicaid as of October 1, 1996.
B2OND6	Number of outpatient non-methadone clients whose primary source of payment was Medicare as of October 1, 1996.
B2OND7	Number of outpatient non-methadone clients whose primary source of payment was other public payment as of October 1, 1996.
B2OND8	Number of outpatient non-methadone clients whose primary source of payment was unknown as of October 1, 1996.
B2ONE1	Number of outpatient non-methadone clients whose principal drug of abuse was heroin/other opiates as of October 1, 1996.
B2ONE2	Number of outpatient non-methadone clients whose principal drug of abuse was cocaine as of October 1, 1996.
B2ONE3	Number of outpatient non-methadone clients whose principal drug of abuse was benzodiazepines as of October 1, 1996.

Item Descriptions

Item	Item description
Items pertaining to outpatient non-methadone facilities (continued)	
B2ONE4	Number of outpatient non-methadone clients whose principal drug of abuse was barbiturates as of October 1, 1996.
B2ONE5	Number of outpatient non-methadone clients whose principal drug of abuse was amphetamines as of October 1, 1996.
B2ONE6	Number of outpatient non-methadone clients whose principal drug of abuse was marijuana/hashish/THC as of October 1, 1996.
B2ONE7	Number of outpatient non-methadone clients whose principal drug of abuse was PCP/LSD as of October 1, 1996.
B2ONE8	Number of outpatient non-methadone clients whose principal drug of abuse was alcohol as of October 1, 1996.
B2ONE9	Number of outpatient non-methadone clients whose principal drug of abuse was other drugs (not alcohol) as of October 1, 1996.
B2ONE10	Number of outpatient non-methadone clients whose principal drug of abuse was unknown as of October 1, 1996.
C2E1	Number of total outpatient non-methadone admissions in the most recent 12-month period.
D13B	Amount of substance abuse revenue generated by outpatient non-methadone care.
D13BPC	Percentage of substance abuse revenue generated by outpatient non-methadone care.
D16C1	Amount of substance abuse costs incurred by outpatient non-methadone care.

APPENDIX C

IMPUTATION RATES

Imputation rates

					Percentage of records imputed by method				
Item	Total respondents	Total imputed	Remaining missing values	Percentage imputed	Using UFDS data (1996-1997)	Using Phase II data	Random regression	Hot-deck	Regression for multi-modality facilities
Items pertaining to all facilties									
A9A1	2,395	5	110[1]	0.21%				0.21%	
A9B1	2,395	5	110[1]	0.21%				0.21%	
A9C1	2,395	5	110[1]	0.21%				0.21%	
A9D1	2,395	5	110[1]	0.21%				0.21%	
A9E1	2,395	5	110[1]	0.21%				0.21%	
A9F1	2,395	5	110[1]	0.21%				0.21%	
A9G1	2,395	5	110[1]	0.21%				0.21%	
A9H1	2,395	5	110[1]	0.21%				0.21%	
A9I1	2,395	4	110[1]	0.17%			0.17%		
A9A2	2,395	4	110[1]	0.17%				0.17%	
A9B2	2,395	4	110[1]	0.17%				0.17%	
A9C2	2,395	4	110[1]	0.17%				0.17%	
A9D2	2,395	4	110[1]	0.17%				0.17%	
A9E2	2,395	4	110[1]	0.17%				0.17%	
A9F2	2,395	4	110[1]	0.17%				0.17%	
A9G2	2,395	4	110[1]	0.17%				0.17%	
A9H2	2,395	4	110[1]	0.17%				0.17%	
A9I2	2,395	4	110[1]	0.17%				0.17%	
A9A3	2,395	4	110[1]	0.17%				0.17%	
A9B3	2,395	4	110[1]	0.17%				0.17%	
A9C3	2,395	4	110[1]	0.17%				0.17%	
A9D3	2,395	4	110[1]	0.17%				0.17%	
A9E3	2,395	4	110[1]	0.17%				0.17%	
A9F3	2,395	4	110[1]	0.17%				0.17%	
A9G3	2,395	4	110[1]	0.17%				0.17%	
A9H3	2,395	4	110[1]	0.17%				0.17%	
A9I3	2,395	4	110[1]	0.17%				0.17%	
B1J2	2,395	0		0.00%					
B1J3	2,395	1		0.04%	0.04%				
B3	2,395	15		0.63%				0.63%	
B4	2,395	25		1.04%				1.04%	
B12A	2,395	9		0.38%				0.38%	
B12B	2,395	0		0.00%					
C2F1	2,395	21		0.88%	0.25%	0.08%	0.38%	0.17%	
C4A	2,395	51		2.13%				2.13%	
C4ANUM	2,395	150		6.26%				6.26%	
C4B	2,395	98		4.09%				4.09%	
C4BNUM	2,395	294		12.28%				12.28%	
D1	2,395	4	8	0.17%	0.17%				
D4	2,395	22	8	0.92%	0.46%			0.46%	
D7	2,395	226		9.44%	2.92%	0.08%	6.43%		
D8A	2,395	165	67	6.89%	3.09%	0.04%	3.76%		
D8B	2,395	165	67	6.89%	3.09%	0.04%	3.76%		
D8C	2,395	164	67	6.85%	3.05%	0.04%	3.76%		
D8D	2,395	165	68	6.89%	3.09%	0.04%	3.76%		
D8E	2,395	164	67	6.85%	3.05%	0.04%	3.76%		
D8F	2,395	165	68	6.89%	3.09%	0.04%	3.76%		
D8G	2,395	164	67	6.85%	3.05%	0.04%	3.76%		
D8H	2,395	164	68	6.85%	3.05%	0.04%	3.76%		
D8I	2,395	164	67	6.85%	3.05%	0.04%	3.76%		
D8APC	2,395	34	68	1.42%	1.42%				
D8BPC	2,395	34	68	1.42%	1.42%				
D8CPC	2,395	33	68	1.38%	1.38%				
D8DPC	2,395	34	68	1.42%	1.42%				
D8EPC	2,395	33	68	1.38%	1.38%				
D8FPC	2,395	34	68	1.42%	1.42%				
D8GPC	2,395	33	68	1.38%	1.38%				
D8HPC	2,395	33	68	1.38%	1.38%				
D8IPC	2,395	33	68	1.38%	1.38%				
D14	2,395	254		10.61%		0.25%	10.31%	0.04%	

[1] The 110 respondents with missing values for items A9A1 - A9I3 reported full time equivalent staff in items A9A4 - A9I4 (not shown).

Imputation rates

Item	Total respondents	Total imputed	Remaining missing values	Percentage imputed	Percentage of records imputed by method				
					Using UFDS data (1996-1997)	Using Phase II data	Random regression	Hot-deck	Regression for multi-modality facilities
Items pertaining to all facilties (continued)									
D15A	2,395	271		11.32%			3.30%	8.02%	
D15B	2,395	269		11.23%			3.55%	7.68%	
D15C	2,395	269		11.23%			3.34%	7.89%	
D15APC	2,395	191		7.97%				7.97%	
D15BPC	2,395	183		7.64%				7.64%	
D15CPC	2,395	189		7.89%				7.89%	
D16D	2,395	163		6.81%		0.04%		6.76%	
Items pertaining to inpatient facilties									
B2INA1	343	3		0.87%				0.87%	
B2INA2	343	3		0.87%				0.87%	
B2INA3	343	3		0.87%				0.87%	
B2INB1	343	3		0.87%				0.87%	
B2INB2	343	3		0.87%				0.87%	
B2INB3	343	3		0.87%				0.87%	
B2INB4	343	3		0.87%				0.87%	
B2INB5	343	3		0.87%				0.87%	
B2INB6	343	3		0.87%				0.87%	
B2INC1	343	3		0.87%				0.87%	
B2INC2	343	3		0.87%				0.87%	
B2INC3	343	3		0.87%				0.87%	
B2INC4	343	3		0.87%				0.87%	
B2INC5	343	3		0.87%				0.87%	
B2INC6	343	3		0.87%				0.87%	
B2IND1	343	3		0.87%				0.87%	
B2IND2	343	3		0.87%				0.87%	
B2IND3	343	3		0.87%				0.87%	
B2IND4	343	3		0.87%				0.87%	
B2IND5	343	3		0.87%				0.87%	
B2IND6	343	3		0.87%				0.87%	
B2IND7	343	3		0.87%				0.87%	
B2IND8	343	3		0.87%				0.87%	
B2INE1	343	3		0.87%				0.87%	
B2INE2	343	3		0.87%				0.87%	
B2INE3	343	3		0.87%				0.87%	
B2INE4	343	3		0.87%				0.87%	
B2INE5	343	3		0.87%				0.87%	
B2INE6	343	3		0.87%				0.87%	
B2INE7	343	3		0.87%				0.87%	
B2INE8	343	3		0.87%				0.87%	
B2INE9	343	3		0.87%				0.87%	
B2INE10	343	3		0.87%				0.87%	
B1A2	343	0		0.00%					
B1B2	343	1		0.29%	0.29%				
B1C2	343	1		0.29%	0.29%				
B1A3	343	1		0.29%	0.29%				
B1B3	343	1		0.29%				0.29%	
B1C3	343	1		0.29%				0.29%	
C2A1	343	4		1.17%			0.29%		0.87%
D12A	343	72		20.99%	3.79%		11.37%		5.83%
D12APC	343	33		9.62%	1.17%		2.62%		5.83%
D16A	343	85		24.78%		0.29%	5.25%	7.87%	11.37%
Items pertaining to residential facilties									
B1D2	598	0		0.00%					
B1E2	598	1		0.17%	0.17%				
B1F2	598	1		0.17%	0.17%				
B1D3	598	0		0.00%					
B1E3	598	1		0.17%	0.17%				
B1F3	598	1		0.17%	0.17%				
B2REA1	598	1		0.17%				0.17%	
B2REA2	598	1		0.17%				0.17%	
B2REA3	598	1		0.17%				0.17%	
B2REB1	598	1		0.17%				0.17%	
B2REB2	598	1		0.17%				0.17%	
B2REB3	598	1		0.17%				0.17%	
B2REB4	598	1		0.17%				0.17%	

Imputation rates

					Percentage of records imputed by method				
Item	Total respondents	Total imputed	Remaining missing values	Percentage imputed	Using UFDS data (1996-1997)	Using Phase II data	Random regression	Hot-deck	Regression for multi-modality facilities
Items pertaining to residential facilties (continued)									
B2REB5	598	1		0.17%				0.17%	
B2REB6	598	1		0.17%				0.17%	
B2REC1	598	1		0.17%				0.17%	
B2REC2	598	1		0.17%				0.17%	
B2REC3	598	1		0.17%				0.17%	
B2REC4	598	1		0.17%				0.17%	
B2REC5	598	1		0.17%				0.17%	
B2REC6	598	1		0.17%				0.17%	
B2RED1	598	1		0.17%				0.17%	
B2RED2	598	1		0.17%				0.17%	
B2RED3	598	1		0.17%				0.17%	
B2RED4	598	1		0.17%				0.17%	
B2RED5	598	1		0.17%				0.17%	
B2RED6	598	1		0.17%				0.17%	
B2RED7	598	1		0.17%				0.17%	
B2RED8	598	1		0.17%				0.17%	
B2REE1	598	1		0.17%				0.17%	
B2REE2	598	1		0.17%				0.17%	
B2REE3	598	1		0.17%				0.17%	
B2REE4	598	1		0.17%				0.17%	
B2REE5	598	1		0.17%				0.17%	
B2REE6	598	1		0.17%				0.17%	
B2REE7	598	1		0.17%				0.17%	
B2REE8	598	1		0.17%				0.17%	
B2REE9	598	1		0.17%				0.17%	
B2REE10	598	1		0.17%				0.17%	
C2B1	598	4		0.67%	0.17%				0.50%
D12B	598	33		5.52%	1.17%		2.34%		2.01%
D12BPC	598	18		3.01%	0.50%		0.50%		2.01%
D16B	598	31		5.18%		0.17%	0.84%	2.01%	2.17%
Items pertaining to all outpatient facilties									
B1G2	1,761	0		0.00%					
B1G3	1,761	0		0.00%					
C2C1	1,761	3		0.17%	0.17%				
C2D1	1,761	4		0.23%	0.06%		0.06%		0.11%
C2E1	1,761	2		0.11%	0.11%				
D12C	1,761	175		9.94%	2.73%	0.11%	5.79%		1.31%
D12CPC	1,761	64		3.63%	0.74%		1.59%		1.31%
D16C	1,761	201		11.41%		0.17%	3.29%	5.34%	2.61%
Items pertaining to outpatient methadone facilities									
B1H2	418	0		0.00%					
B1H3	418	0		0.00%					
B2OMA1	418	1		0.24%				0.24%	
B2OMA2	418	1		0.24%				0.24%	
B2OMA3	418	1		0.24%				0.24%	
B2OMB1	418	1		0.24%				0.24%	
B2OMB2	418	1		0.24%				0.24%	
B2OMB3	418	1		0.24%				0.24%	
B2OMB4	418	1		0.24%				0.24%	
B2OMB5	418	1		0.24%				0.24%	
B2OMB6	418	1		0.24%				0.24%	
B2OMC1	418	1		0.24%				0.24%	
B2OMC2	418	1		0.24%				0.24%	
B2OMC3	418	1		0.24%				0.24%	
B2OMC4	418	1		0.24%				0.24%	
B2OMC5	418	1		0.24%				0.24%	
B2OMC6	418	1		0.24%				0.24%	
B2OMD1	418	1		0.24%				0.24%	
B2OMD2	418	1		0.24%				0.24%	
B2OMD3	418	1		0.24%				0.24%	
B2OMD4	418	1		0.24%				0.24%	
B2OMD5	418	1		0.24%				0.24%	
B2OMD6	418	1		0.24%				0.24%	
B2OMD7	418	1		0.24%				0.24%	
B2OMD8	418	1		0.24%				0.24%	

Imputation rates

Item	Total respondents	Total imputed	Remaining missing values	Percentage imputed	Percentage of records imputed by method				
					Using UFDS data (1996-1997)	Using Phase II data	Random regression	Hot-deck	Regression for multi-modality facilities
Items pertaining to outpatient methadone facilities (continued)									
B2OME1	418	1		0.24%				0.24%	
B2OME2	418	1		0.24%				0.24%	
B2OME3	418	1		0.24%				0.24%	
B2OME4	418	1		0.24%				0.24%	
B2OME5	418	1		0.24%				0.24%	
B2OME6	418	1		0.24%				0.24%	
B2OME7	418	1		0.24%				0.24%	
B2OME8	418	1		0.24%				0.24%	
B2OME9	418	1		0.24%				0.24%	
B2OME10	418	1		0.24%				0.24%	
C2D1	418	4		0.96%	0.24%		0.48%		0.24%
D13A	418	42		10.05%	2.39%		5.02%		2.63%
D13APC	418	22		5.26%	0.72%		1.91%		2.63%
D16C2	418	49		11.72%		0.24%	2.15%	5.98%	3.35%
Items pertaining to outpatient non-methadone facilities									
B1I2	1,761	0		0.00%					
B1I3	1,761	0		0.00%					
B2ONA1	1,435	2		0.14%				0.14%	
B2ONA2	1,435	2		0.14%				0.14%	
B2ONA3	1,435	2		0.14%				0.14%	
B2ONB1	1,435	2		0.14%				0.14%	
B2ONB2	1,435	2		0.14%				0.14%	
B2ONB3	1,435	2		0.14%				0.14%	
B2ONB4	1,435	2		0.14%				0.14%	
B2ONB5	1,435	2		0.14%				0.14%	
B2ONB6	1,435	2		0.14%				0.14%	
B2ONC1	1,435	2		0.14%				0.14%	
B2ONC2	1,435	2		0.14%				0.14%	
B2ONC3	1,435	2		0.14%				0.14%	
B2ONC4	1,435	2		0.14%				0.14%	
B2ONC5	1,435	2		0.14%				0.14%	
B2ONC6	1,435	2		0.14%				0.14%	
B2OND1	1,435	2		0.14%				0.14%	
B2OND2	1,435	2		0.14%				0.14%	
B2OND3	1,435	2		0.14%				0.14%	
B2OND4	1,435	2		0.14%				0.14%	
B2OND5	1,435	2		0.14%				0.14%	
B2OND6	1,435	2		0.14%				0.14%	
B2OND7	1,435	2		0.14%				0.14%	
B2OND8	1,435	2		0.14%				0.14%	
B2ONE1	1,435	2		0.14%				0.14%	
B2ONE2	1,435	2		0.14%				0.14%	
B2ONE3	1,435	2		0.14%				0.14%	
B2ONE4	1,435	2		0.14%				0.14%	
B2ONE5	1,435	2		0.14%				0.14%	
B2ONE6	1,435	2		0.14%				0.14%	
B2ONE7	1,435	2		0.14%				0.14%	
B2ONE8	1,435	2		0.14%				0.14%	
B2ONE9	1,435	2		0.14%				0.14%	
B2ONE10	1,435	2		0.14%				0.14%	
C2E1	1,435	2		0.14%	0.14%				
D13B	1,435	150		10.45%	2.65%	0.14%	5.57%		2.09%
D13BPC	1,435	58		4.04%	0.56%	0.07%	1.32%		2.09%
D16C1	1,435	173		12.06%		0.21%	3.34%	4.67%	3.83%

APPENDIX D

IMPUTATION PROCESS VARIABLE DESCRIPTIONS

Hot Deck Variables

The categorical variable for clients uses Brandeis' 4-level analytical variable, and creates two categories out of the category with the largest client values.

```
If B1J2 <= 16 then BUCLNT2 = 1;
else if B1J2 <= 40 then BUCLNT2 = 2;
else if B1J2 <= 100 then BUCLNT2 = 3;
else if B1J2 <= 224 then BUCLNT2 = 4; and
else BUCLNT2 = 5,
```

A 3-level type of ownership variable was created.

```
If _A6 = 1 then OWN = 1;
else if _A6 = 2 then OWN = 2; and
else OWN = 3.
```

The following boundary variables were created to be used to target records with missing values in the 'B' block of items, so that it will result in creating less hard boundary cells.

```
Let B_BC    = 1, if B1B1        = 1 and B1C1 = 1; and
                                = 0, otherwise.

Let B_HBOUND      = 1, if B1A1 = 1 and B1I1 = 1;
                  = 2, if B1A1 = 1;
                  = 3, if B1D1 = 1;
                  = 4, if B1H1 = 1;
                  = 5, if B1I1 = 1; and
                  = 0, otherwise.
```

In order to impute for records that have missing B12B when B1J3 > 0, boundary variable, B_MM, was created as follows:

```
If B1J3 > 0 then set B_MM =1; and
Else set B_MM = 0.
```

In order to have the software treat different missing value codes the same, the following variable was created:

```
B_D1        =       if D1 = any missing or .S; and
            =       D1, otherwise.
```

In order to protect from imputing clients admitted who are pregnant for facilities that are male only, a positive imputed value was allowed for pregnant females admitted only for facilities that had female clients on Oct. 1, 1996. Therefore, the boundary variable C_FEM, was created, where,

C_FEM = 1, if B2INA2 or B2REA2 or B2OMA2 or B2ONA2 > 0; and
= 0, otherwise.

A 5-level variable CTC2F1 was created from total admissions. Five levels were made in order so that a relatively high proportion from a donor that is smaller than the donee, is not applied to the donee, which would result in high numbers for C4BNUM.

If 0 <= C2F1 <= 25th percentile, then CTC2F1 = 1;
Else if C2F1 <= 50th percentile, then CTC2F1 = 2;
Else if C2F1 <= 75th percentile, then CTC2F1 = 3;
Else if C2F1 <= 90th percentile, then CTC2F1 = 4; and
Else CTC2F1 = 5.

A 4-level variable CTA9I1 was created from total full time staff as:

If 0 <= A9I1 <= 25th percentile, then CTA9I1 = 1;
Else if A9I1 <= 50th percentile, then CTA9I1 = 2;
Else if A9I1 <= 75th percentile, then CTA9I1 = 3; and
Else CTA9I1 = 4.

A 4-level variable CTD7 was created from total revenues as:

If 0 <= D7 <= 25th percentile, then CTD7 = 1;
Else if D7 <= 50th percentile, then CTD7 = 2;
Else if D7 <= 75th percentile, then CTD7 = 3; and
Else CTD7 = 4.

Regression dummy variables

If _A6 = 1 then OWN1 = 1;
Else OWN1 = 0;
If _A6 = 2 then OWN2 = 1; and
Else OWN2 = 0.

If PSUTYPE2 = 1 then PSUT1 = 1;
Else PSUT1 = 0;
If PSUTYPE2 = 2 then PSUT2 = 1; and
Else PSUT2 = 0.

If CENREG = 1 then CREG1 = 1;
Else CREG1 = 0;
If CENREG = 2 then CREG2 = 1;
Else CREG2 = 0;
If CENREG = 3 then CREG3 = 1; and
Else CREG3 = 0.

Random Regression Continuous Variables

LOGB1 = natural log of B1J2,
LOGC2 = natural log of C2F1
LOGA9 = natural log of A9I1
LOGREV = natural log of D7 after D7 was converted to represent substance abuse only

Regression Continuous Variables

PB1A2 = B1A2 / B1J2;
PB1D2 = B1D2 / B1J2;
PB1A2_2 = B1A2 / (B1J2 – B1D2);
PB1A2_3 = B1A2 / (B1J2 – B1G2);
PB1D2_2 = B1D2 / (B1J2 – B1A2);
PB1H2 = B1H2 / B1G2;

PC2A1 = C2A1 / C2F1;
PC2B1 = C2B1 / C2F1;
PC2A1_2 = C2A1 / (C2F1 – C2B1);
PC2B1_2 = C2B1 / (C2F1 – C2A1);
PC2A1_3 = C2A1 / (C2F1 – C2C1);
PC2D1 = C2D1 / C2F1;

PD12A = D12A / D12D;
PD12B = D12B / D12D;
PD12A_3 = D12A / (D12D – D12C);
PD13A = D13A / D13C;

PD16A = D16A / (D16E – D16D);
PD16B = D16B / (D16E – D16D); and
PD16C2 = D16C2 / D16C.