2. SAMPLE DESIGN

This chapter documents the procedures used to select laboratories for the National Inventory of Clinical Laboratory Testing Services. The sample was designed to allow CDC to determine the proportions of laboratories measuring specific analytes, performing specific tests, and performing specific combinations of analytes, test systems, and biological specimens.

The sample for the NICLTS was selected in several steps. First, an initial sample of 6,000 laboratories was selected. Next, a subsample of 2,503 laboratories was selected for the primary survey sample. This sample was divided into two parts, corresponding to the Phase I and II samples. Laboratories holding moderate and high complexity CLIA certificates were fielded in Phase I and tabulated on site. Laboratories holding waived or Provider Performed Microscopy (PPM) CLIA certificates were surveyed in Phase II. This second part of the survey employed a mail-telephone methodology where laboratories were mailed a short data collection form and then contacted by telephone by a trained data tabulator to assist with the completion of the form and collect the data.

2.1 Sampling Frame

The sampling frame consisted of 157,779 records in the July 1996 OSCAR database maintained by the Department of Health and Human Services. Each record corresponded to a CLIA ID number in the PROVNUM field. The frame was stratified by the 10 DHHS designated geographic regions and six laboratory groups based on the 23 Health Care Financing Administration clinical laboratory types. To simplify the survey, territories and Commonwealths of the United States were excluded from the geographic regions. These laboratories were (1) located in the 50 states or the District of Columbia and (2) had completed a Form HCFA-116, Clinical Laboratory Application. Table 2-1 shows a distribution of these laboratories by geographic region and by the six major laboratory groupings.

Prior to sampling, the frame was stratified by physician office laboratory category (POL vs. non-POL), application category (moderate/high complexity testing vs. waived/PPM testing), then sorted by region (ten

Table2-1. Distribution of laboratories on the sampling frame by region and laboratory group

	Laboratory Group						
Region	POL	Other Ambulatory	Hospice, Nursing home	Hospital	Independent, Blood bank	Specialty	Total
1. Northeast	4,819	886	1,143	434	346	893	8,521
2. New York, New Jersey	10,037	724	708	557	620	1,412	14,058
3. Mid-Atlantic	9,666	1,244	1,306	829	572	2,262	15,879
4. Southeast	17,941	2,839	2,240	1,824	1,080	4,508	30,432
5. Midwest (North)	16,209	2,624	3,398	1,513	765	4,708	29,217
6. South (Central)	9,764	3,915	2,081	1,378	734	3,181	21,053
7. Midwest (Central)	4,033	957	1,531	602	207	1,654	8,984
8. Mountain	2,313	730	536	390	176	826	4,971
9. West	11,931	2,171	1,369	980	884	2,144	19,479
10. Northwest	2,830	496	465	295	212	887	5,185
Total	89,543	16,586	14,777	8,802	5,596	22,475	157,779

levels) and laboratory type within group (see Tables 2-2, 2-3, and 2-4). This procedure helped to ensure the selection of laboratories with a variety of characteristics. While the main focus of the study was on laboratory group (rather than type), sorting by laboratory type within group marginally improved the sample distribution by laboratory type without changing the expected sample sizes by laboratory group. Region and laboratory type were taken from the OSCAR database. Application category was constructed from the CLIA application type. POL category and laboratory group were constructed from laboratory type. Laboratory group had six levels, separating POLs from other ambulatory sites. This allowed the sample size of POLs to be monitored more closely.

2.2 Sample Design Parameters

The sample was designed to allow CDC to determine what proportion of laboratories in various categories measured specific analytes; what proportion perform specific tests; and what proportion perform specific combinations of analytes, test systems, and biological specimens. No attempt was made to ensure that every rare analyte would be 'included in the sampled data. Although the study design included the collection of volume data, the sample was not optimally designed to provide national estimates of test volume.

Because of the large proportion of POLs in the sampling frame, POLs were sampled at one-half the rate of other laboratory groups. This subsampling led to substantial cost savings with only a small loss in the precision of estimates for the POL laboratory group.

2.3 Sample Selection

To meet the desired precision requirements, the study design called for 1,834 participating laboratories. Westat first selected a sample of 6,000 laboratories. This large initial sample was then subsampled to a primary sample of 2,503 as described below. The increase from 1,834 to 2,503 allowed for nonresponse. The balance of the initial sample of 6,000 that was not in the primary sample constituted the reserve sample. The reserve sample allowed a substantial margin to offset the possible losses in the primary sample from larger than expected nonresponse, laboratories no longer in business, and possible duplicate records on the frame. The initial sample of 6,000 laboratories was selected with equal probability using systematic random sampling. Table 2-5 shows the

Table 2-2. Definition of regions

	Region	State
1.	Northwest	CT, ME, MA, NH, RI, VT
2.	New York, New Jersey	NJ, NY
3.	Mid-Atlantic	DE, DC, MD, PA, VA, WV
4.	Southeast	AL, FL, GA, KY, MS, NC, SC, TN
5.	Midwest (North)	IL, IN, MI, MN, OH, WI
6.	South (Central)	AR, LA, NM, TX, OK
7.	Midwest (Central)	IA, KS, MO, NE
8.	Mountain	CO, MT, ND, SD, UT, WY
9.	West	CA, HI, NV, AZ
10.	Northwest	AK, ID, OR, WA

Table 2-3. Definition of laboratory groups

Laboratory Group	Laboratory Type
1. Physicians Office Laboratory (POL)	Physician office
2. Other Ambulatory	Community clinic Home health agency Student health service Health Maintenance Organization (HMO)
3. Hospice/Nursing Home	Hospice Skilled nursing/nursing facility
4. Hospital	Hospital
5. Independent/Blood Bank	Independent Tissue bank/repository Blood bank
6. Specialty	Ambulatory surgery center Comprehensive outpatient rehabilitation Ancillary test site End stage renal disease dialysis Health fair Industrial Insurance Intermediate care facility mental retarded Mobile unit Pharmacy Other practitioner Other

Table 2-4. Definition of application categories

Application category	Application type (type of CLIA certificate)
1. Waived/PPM	Waived Microscopy
2. Moderate/High complexity	Certificate Accreditation Partial accreditation

distribution of these 6,000 laboratories by region and laboratory group.

The primary sample of 2,503 laboratories was selected from the initial sample of 6,000 using stratified systematic sampling. In stratified systematic sampling, every n_s-*th* laboratory is sampled from a sorted file, where n_s is set for each stratum *s* so that the target sample size is achieved for each stratum. Four sampling strata were formed by cross-classifying POL category (POL versus not POL) and application category. Table 2-6 shows the distribution of the final sample by region and laboratory group. Table 2-7 shows the distribution of the 2,503 laboratories by the POL and application strata.

Release of Reserve Samples

In both Phase I and Phase II, additional laboratories were released in controlled groups from the respective reserve samples. The release took place after enrollment had been attempted with a sufficient number of the primary sample laboratories to make projections of final enrollment rates.

To form the release groups, the reserve sample for each phase was sorted by region, group, and laboratory type. Within these categories, facilities were sorted at random and then the entire file was systematically divided into four release groups of approximately equal size. Each resulting release group constituted a representative subsample, so that adding the release groups to the primary sample did not invalidate the representative nature of the overall sample.

Table 2-5. Distribution of laboratories in the initial sample by region and laboratory group

	Laboratory Group						
Region	POL	Other Ambulatory	Hospice, Nursing home	Hospital	Independent, Blood bank	Specialty	Total
1. Northeast	184	34	43	17	12	35	325
2. New York, New Jersey	382	27	27	21	23	55	535
3. Mid-Atlantic	367	46	51	31	21	87	603
4. Southeast	682	108	85	69	42	171	1,157
5. Midwest (North)	617	100	129	57	30	178	1,111
6. South (Central)	371	149	79	53	28	121	801
7. Midwest (Central)	154	37	57	23	8	63	342
8. Mountain	88	28	20	15	7	31	189
9. West	453	82	52	38	34	82	741
10. Northwest	108	18	18	11	8	33	196
Total	3,406	629	561	335	213	856	6,000

Table 2-6. Distribution of laboratories in the primary sample by region and laboratory group

	Laboratory Group						
Region	POL	Other Ambulatory	Hospice, Nursing home	Hospital	Independent, Blood bank	Specialty	Total
1. Northeast	54	20	26	9	6	21	136
2. New York, New Jersey	109	16	17	11	13	32	198
3. Mid-Atlantic	107	28	30	17	12	50	244
4. Southeast	196	65	51	38	23	100	473
5. Midwest (North)	181	59	78	31	17	105	471
6. South (Central)	107	90	47	30	14	72	360
7. Midwest (Central)	44	21	35	13	4	37	154
8. Mountain	25	16	13	8	4	18	84
9. West	133	48	32	21	19	47	300
10. Northwest	32	11	11	6	4	19	83
Total	988	374	340	184	116	501	2,503

Table 2-7. Distribution of laboratories in the initial, primary, and reserve samples by stratum

7a. Initial sample:

Stratum	POL category	Application category	Number of laboratories
1	POL	Waived/PPM	2,089
2	POL	Moderate/high complexity	1,317
3	non POL	Waived/PPM	1,715
4	non POL	Moderate/high complexity	879
Total			6,000

7b. Primary sample:

Stratum	POL category	Application category	Number of laboratories
1	POL	Waived/PPM	637
2	POL	Moderate/high complexity	351
3	non POL	Waived/PPM	1,046
4	non POL	Moderate/high complexity	469
Total			2,503

7c. Reserve sample:

Stratum	POL category	Application category	Number of laboratories
1	POL	Waived/PPM	1,452
2	POL	Moderate/high complexity	966
3	non POL	Waived/PPM	669
4	non POL	Moderate/high complexity	410
Total			3,497

Phase I Sample

For Phase 1, a primary sample of 820 laboratories and a reserve sample of 110 laboratories were contacted for enrollment, for a total released sample of 930 laboratories. Table 2-8 shows the distribution of the Phase I primary sample by region and laboratory group. Table 2-9 shows the Phase I reserve sample.

Phase II Sample

For Phase II a primary sample of 1,683 laboratories and a reserve sample of 176 laboratories were contacted for enrollment, for a total released sample of 1,859 laboratories. The distribution of the waived/PPM laboratories on the OSCAR database is shown in Table 2-10. Table 2-11 shows the distribution of the Phase II primary sample by region and laboratory group. Table 2-12 shows the Phase II reserve sample.

2.4 Other Sampling Issues

Several other sampling issues affected the study. These are subsampling laboratories with multiple locations, and matching laboratories with CLIA ID numbers. Each of these is discussed in detail in the sections that follow.

Subsampling

During Phase I, subsampling was used in two situations. The first of these was that nursing stations were subsampled in hospitals, HMOs and nursing homes; the second was that daily logbooks were subsampled wherever encountered.

At hospitals, HMOs, and nursing homes, tabulators were instructed to enter a list of all nursing stations, and the Tabulation Device randomly selected the stations to tabulate. First, the tabulator identified "homogeneous" nursing stations, meaning that the same tests were performed at all stations. For example, nursing

Table 2-8. Distribution of Phase I primary sample by region and laboratory group

Region	Laboratory Group						
	POL	Other Ambulatory	Hospice, Nursing home	Hospital	Independent, Blood bank	Specialty	Total
1. Northeast	20	7	0	7	5	7	46
2. New York, New Jersey	39	4	1	10	7	10	71
3. Mid-Atlantic	39	6	0	15	9	17	86
4. Southeast	79	11	0	32	17	32	171
5. Midwest (North)	59	11	0	26	12	30	138
6. South (Central)	38	9	0	24	11	20	102
7. Midwest (Central)	17	4	0	12	3	10	46
8. Mountain	10	5	0	7	3	6	31
9. West	38	10	0	18	16	16	98
10. Northwest	12	4	0	5	3	7	31
Total	351	71	1	156	86	155	820

Table 2-9. Distribution of Phase I reserve sample by region and laboratory group

	Laboratory Group						
Region	POL	Other Ambulatory	Hospice, Nursing home	Hospital	Independent, Blood bank	Specialty	Total
1. Northeast	53	5	0	7	5	5	75
2. New York, New Jersey	111	4	0	9	7	9	140
3. Mid-Atlantic	106	4	1	13	7	15	146
4. Southeast	218	9	1	28	15	27	298
5. Midwest (North)	160	10	1	23	10	25	229
6. South (Central)	107	8	1	20	11	17	164
7. Midwest (Central)	47	4	0	9	4	9	73
8. Mountain	28	4	0	6	2	6	46
9. West	103	9	0	15	13	15	155
10. Northwest	33	3	0	4	4	6	50
Total	966	60	4	134	78	134	1,376

Table 2- 10. Distribution of waived and PPM laboratories on the OSCAR database by region and laboratory group

	Laboratory Group						
Region	POL	Other Ambulatory	Hospice, Nursing home	Hospital	Independent, Blood bank	Specialty	Total
1. Northeast	2,917	587	1,137	74	61	597	5,373
2. New York, New Jersey	6,083	496	694	42	255	933	8,503
3. Mid-Atlantic	5,847	959	1,295	88	143	1,439	9,771
4. Southeast	10,128	2,312	2,210	232	252	2,950	18,084
5. Midwest (North)	10,471	2,085	3,365	216	200	3,240	19,577
6. South (Central)	5,946	3,469	2,062	220	167	2,200	14,064
7. Midwest (Central)	2,345	757	1,522	54	22	1,142	5,842
8. Mountain	1,320	512	533	51	30	515	2,961
9. West	8,221	1,665	1,362	123	129	1,335	12,835
10. Northwest	1,656	312	463	47		542	3,042
Total	54,934	13,154	14,643	1,147	1,281	14,893	100,052

Table 2-11. Distribution of Phase II primary sample by region and laboratory group

	Laboratory Group						
Region	POL	Other Ambulatory	Hospice, Nursing home	Hospital	Independent, Blood bank	Specialty	Total
1. Northeast	34	13	26	2	1	14	90
2. New York, New Jersey	70	12	16	1	6	22	127
3. Mid-Atlantic	68	22	30	2	3	33	158
4. Southeast	117	54	51	6	6	68	302
5. Midwest (North)	122	48	78	5	5	75	333
6. South (Central)	69	81	47	6	3	52	258
7. Midwest (Central)	27	17	35	1	1	27	108
8. Mountain	15	11	13	1	1	12	53
9. West	95	38	32	3	3	31	202
10. Northwest	20	7	11	1	1	12	52
Total	637	303	339	28	30	346	1,683

Table 2-12. Distribution of Phase II reserve sample by region and laboratory group

	Laboratory Group						
Region	POL	Other Ambulatory	Hospice, Nursing home	Hospital	Independent, Blood bank	Specialty	Total
1. Northeast	77	9	17	1	1	9	114
2. New York, New Jersey	162	7	10	1	3	14	197
3. Mid-Atlantic	154	14	20	1	2	22	213
4. Southeast	268	34	33	3	4	44	386
5. Midwest (North)	276	31	50	3	3	48	411
6. South (Central)	157	51	31	3	3	32	277
7. Midwest (Central)	63	12	22	1	0	17	115
8. Mountain	35	8	7	1	1	7	59
9. West	217	25	20	2	2	20	286
10. Northwest	43	4	7	1	0	8	63
Total	1,452	195	217	17	19	221	2,121

stations at one nursing home might all perform the same simple test menu such as measuring glucose levels with a single analyte instrument and performing fecal occult blood tests using a single manufacturer's test system. Then, two homogeneous nursing stations were subsampled per location. The random sampling procedure was self contained in the Tabulation Device. All nursing stations that were not homogeneous were fully tabulated.[3]

In sampling daily logbooks, 20 days per log were selected. As with nursing stations, a subsampling module was created in the Tabulation Device to select days from logbooks. The selection algorithm used systematic sampling with a random start date. The tabulator activated this module as necessary. The tabulator entered the days of the week that the laboratory was open for business and the number of holidays observed in 1996, and the Tabulation Device selected a sample of 20 dates for the tabulator to look up in the logbook. The Tabulation Device used the number of days that a laboratory was open in calendar year 1996 to calculate the number of business days to use as the interval between selected days.

Subsampling was not used at all during Phase II of the survey.

Laboratories with Multiple Locations

Many laboratories operated in several locations under the same CLIA ID number. Such laboratories were encountered both in Phase I and PhaseII. In most cases, the individual locations were treated as separate laboratories for purposes of data collection. For example, in Phase I, tabulators were sent to each location and a separate data collection site visit was conducted for each.

Laboratories that operated in four or more locations were evaluated for possible subsampling. In most of these cases, Westat selected a sample of the locations for tabulation. During the data processing, an adjustment was made to the estimates to reflect this procedure. Subsampling of locations occurred only in Phase I, where there was a large cost associated with visiting each location. In Phase II, the tabulators were able to obtain data from all locations of cooperating sampled laboratories.

3 In theory, a hospital, HMO, or nursing home could have two or more sets of nursing stations with each set performing a well-defined but distinct list of tests. In practice, however, this situation occurred only once. The largest set of nursing stations was subsampled and the remaining nursing stations were fully tabulated.

The largest number of locations under one CLIA ID number encountered during Phase I was 80, 20 of which were subsampled for tabulation. The largest number encountered in Phase II was 228, but one contact was able to provide the needed information for all locations.

Identification of CLIA ID Numbers

Part of the data collection protocol was to confirm the CLIA ID number. While the laboratory's reported CLIA ID number and the sampled CLIA ID number usually matched, the laboratory was occasionally unable to find the number or, more rarely, its reported CLIA ID number did not match the sampled CLIA ID number, or it had several CLIA ID numbers. Sections 5.1.1 and 5.2.1 discuss how Westat resolved these cases in Phases I and II, respectively.

The underlying principle was that the laboratory was sampled by selecting the CLIA ID number on the OSCAR database. In the case of unconfirmed or nonmatching numbers, if the address and name of the laboratory matched, or a chain of evidence existed that linked a new laboratory name or address to the one associated with the sampled CLIA ID number in OSCAR, it was assumed that the laboratory located should be included in the sample.

The case of multiple CLIA ID numbers presented a more complex problem. Westat attempted to identify the testing performed for the sampled CLIA ID number and to tabulate only that testing. Where this was not possible, data were tabulated for all CLIA ID numbers and a record of the numbers was kept.

Multiple CLIA ID Numbers on the OSCAR Database

Where the CLIA ID number at the laboratory did not match the sampled number, all numbers were recorded and referred back to the home office. Some of these numbers matched other CLIA ID numbers in the OSCAR database and turned out to be affiliated with the same laboratory. It is not clear if these numbers resulted from multiple certification applications. In any case, the presence of multiple CLIA ID numbers for the same laboratory presents a sampling problem often encountered in establishment surveys. Some

establishments-typically larger ones- may have multiple listings and thus multiple chances of selection. In extreme cases, this problem can result in biased estimates.

One correction for the problem is to identify all multiple certificate holders on the sampling frame and assign a single identifying number to all certificates belonging to the same holder. Another correction is to identify all such laboratories only among sampled establishments. The latter solution is usually simpler but must be built into the survey protocol at the start of the study; CDC and Westat made a joint decision early in the NICLTS not to allow for such multiple listings in the on-site protocol. Because the number of cases was small, it is unlikely that this problem has affected the survey estimates from the NICLTS. However, this problem should be considered carefully in any future survey work using the OSCAR database as a sampling frame.