A Two-Stage Sampling Method for Clinical Surveillance of Individuals in Care for HIV Infection in the United States

Patrick S. Sullivan, DVM, PhD[a]
John M. Karon, PhD[a]
Faye E. Malitz, MPH[b]
Stephanie Broyles, MPH[c]
Eve D. Mokotoff, MPH[d]
Susan E. Buskin, PhD[e]
Patricia L. Fleming, PhD[a]

SYNOPSIS

Objectives. The goals of this study were two-fold: (*1*) to describe methods for drawing a population-based sample of individuals in care for HIV infection and (*2*) to compare data from the sample with data from existing surveillance systems that describe care for HIV.

Methods. The authors implemented a two-stage sampling method, using local HIV/AIDS surveillance data as a sampling frame of HIV care providers in three states. At selected providers, medical records of a random sample of patients were abstracted.

Results. The medical records of a number of patients, ranging from 253 to 374 individuals per state, were abstracted. The demographics of sampled individuals and of individuals reported to the local HIV/AIDS surveillance program were similar; however, differences existed in the proportion of individuals receiving HIV care consistent with treatment guidelines between the sample and a contemporary facility-based supplemental surveillance project. The median design effect for outcomes collected in the sample was 1.8 (range=0.5–29.6).

Conclusions. This survey method is feasible for collecting population-based data on patients in care for HIV. Sample size and some design elements should be changed in future studies to increase precision of estimates and usefulness of data for local planning and evaluation.

[a]Division of HIV/AIDS Prevention—Surveillance and Epidemiology, National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA

[b]Office of Science and Epidemiology, HIV/AIDS Bureau, Health Resources and Services Administration, Rockville, MD

[c]HIV/AIDS Surveillance Program, Louisiana Office of Public Health, New Orleans, LA

[d]HIV/AIDS Surveillance, Michigan Department of Community Health, Detroit and Lansing, MI

[e]Public Health–Seattle & King County; Department of Epidemiology, University of Washington, Seattle, WA

Address correspondence to: Patrick S. Sullivan, DVM, PhD, Centers for Disease Control and Prevention, 1600 Clifton Rd. NE, MS E46, Atlanta, GA 30333; tel. 404-420-2267; fax 404-639-8640; e-mail <pss0@cdc.gov>.

HIV infection is a major public health problem in the United States, and billions of dollars of public funds are spent annually on the care and treatment of people infected with HIV. Monitoring the morbidity of individuals living with HIV infection is important to (*1*) understand the population impact of advances in the medical management of HIV infection; (*2*) document needs for and utilization of care services; (*3*) inform guidelines for monitoring clinical status, care, and treatment; and (*4*) document the impact of public funds spent on care for people living with HIV infection.[1–3]

Health departments are primarily responsible for administering and evaluating federal, state, and local funds for care. Although AIDS surveillance data have represented a nearly complete census of all individuals with AIDS and are available in all health departments, these data are inadequate to meet the states' information needs for planning HIV-related services. Surveillance case reports contain only limited demographic, behavioral risk, and clinical data; supplemental studies are required to obtain data on access to, use of, quality of, and impact of treatment.

Ideally, data on clinical outcomes of HIV care should be representative and useful at both the national and local levels. In the past, different methods of data collection have had different strengths and weaknesses. During the mid-1990s, the HIV Cost and Services Utilization Study (HCSUS) estimated HIV health care costs and utilization in a nationally representative sample of HIV facilities and patients.[4,5] This watershed survey produced useful national-level estimates of HIV care utilization, but it did not allow for calculation of state level estimates for local planning purposes. In the late 1980s, the Centers for Disease Control and Prevention (CDC) and 10 state and local health departments developed the Adult and Adolescent Spectrum of HIV-related Disease Study (ASD), an ongoing longitudinal review of medical records of HIV-infected individuals in more than 100 hospitals and clinics chosen by the participating health departments.[6,7] ASD has collected observational data on morbidity, treatment, insurance, and health services utilization.[8] However, the representativeness of ASD is limited and inference to the entire United States is not possible.

In an attempt to develop a surveillance method for clinical outcomes that would both provide representative data and be useful at the local level, the CDC and the Health Resources and Services Administration instituted a pilot project in 1998: the Survey of HIV Disease and Care Project (SHDC). The primary goal of the pilot study was to develop and assess the feasibility of a two-stage sampling design to produce a representative sample of patients in care for HIV infection in the U.S., using HIV and AIDS case subject reporting data as a sampling frame of HIV care providers. The goal of this report is to describe the method and its feasibility and to compare the data from SHDC with other existing systems that measure characteristics of individuals in care for HIV infection. We make recommendations for the design of future surveillance activities at the state and local levels that will provide representative data for targeting and evaluating programs at modest incremental cost. Finally, we describe plans for an approach with similar methodology that can generate nationally representative data to provide a sound basis for allocation of resources for HIV prevention, treatment, and care.

METHODS

Survey of HIV Disease and Care Project

Participating areas. The survey was conducted in three project areas during 1999, to collect information about care for HIV infection delivered during 1998. The project areas were King County in Washington State, which includes the Seattle metropolitan area; the Louisiana State Health Regions 1, 2, 3, 4, and 9, representing approximately the southern one-third of Louisiana and including New Orleans and Baton Rouge; and the entire state of Michigan.[9] During the study period, HIV infection without AIDS was reportable to the health departments in Michigan and Louisiana, but not in King County.

Patients were chosen for the survey using a two-stage sampling procedure. The first stage was selection of a sample of medical care providers; the second stage was a weighted, stratified systematic sample of patients within selected providers.

Selection of the sample of providers. The sampling frame for the first stage was based on data from the HIV/AIDS Reporting System (HARS). A medical provider was recorded for all individuals with HIV infection (with or without AIDS) who (*1*) were reported to the state HIV/AIDS surveillance system, (*2*) resided in the eligible geographic area at the time of diagnosis, and (*3*) were presumed to be alive at some time during the study period, 1998. The provider was either the most recent provider reporting the case of HIV infection or AIDS, or, if available, the medical care provider who most recently had ordered CD4 count or viral load tests for the patient (which we presumed to indicate provision of clinical care). The sampling frame of providers consisted of all medical care providers who cared for or reported at least one such case subject (except in Michigan, where only providers who cared for or reported at least three case subjects were included) to the state or county health department HIV/AIDS surveillance system. The initial list of providers was reviewed to remove providers who had closed or merged, who were sources of HIV infection reports that did not provide medical care (such as HIV testing sites that did not provide primary HIV care), and who were outside the specified geographic area. Correctional facilities were excluded from the provider sampling frame in all sites because of the difficulty in obtaining patient lists and medical records from these facilities.

For each area, the sampling frame of providers was stratified as small, medium, or large, based on numbers of patients attributed to that provider in the HIV/AIDS surveillance system. The definitions of *small, medium,* and *large* providers were different in all three areas: King County stratified the sample as small <30 patients, medium 30–99 patients, and large ≥ 100; Louisiana stratified the sample as small <20, medium 20–100, and large >100; and Michigan stratified the sample as small <10, medium 10–99, and large ≥ 100. In all sites, large providers were stratified on Ryan

White Comprehensive AIDS Resources Emergency (CARE) Act funding to ensure that adequate numbers of patients in care supported by Ryan White CARE Act (RWCA) funds were included.[10] In Michigan, providers were further stratified by size of the metropolitan statistical area (MSA) in which the providers were located, so that we could compare care by large providers with and without RWCA funds.

Within each stratum, providers were sampled without replacement with probability proportional to the number of HIV/AIDS patients assigned to the provider. To select a provider sample for our study, we generated a list ordered by size within each stratum using the S-Plus sample function, specifying the probability of sampling as the proportion of patients from the stratum at the provider.[11] We obtained the sample by choosing the desired number of providers, starting at the top of the list for each stratum.

Each selected provider was contacted to verify willingness and ability to participate. Participation required that (*1*) the provider could supply a list of all HIV-infected patients in care for the calendar year 1998, including the race and sex of each patient, and (*2*) the provider would allow information to be abstracted from medical records. If a selected provider refused or was unable to participate (e.g., because of inability to supply a list of individuals with HIV infection in care during the year), or if the provider was found later not to be eligible (e.g., because of providing no medical care in the reference year), the next provider on the ordered sample list was approached as an alternate.

Selection of the sample of patients. Within each geographic area, the health department decided how to stratify the patient population on race/ethnicity and sex, based on the demographic characteristics of the HIV-infected population. For example, Louisiana identified strata of black men, "other" men, and women. Within each provider stratum, we attempted to allocate the patient sample equally across the race/sex strata.

We defined a *sampling fraction* for each race/sex stratum based on the total number of patients to be sampled for that race/sex stratum and the total number of patients who had at least one visit to the providers in that provider stratum. For each race/sex stratum, the same sampling fraction was used at all providers within a provider stratum.

Each provider was instructed to develop a list of patients ordered by date of birth or medical record number for each race/sex combination or to provide a list of patients that included the demographics, which would allow health department personnel to create the patient sampling frame. Patients then were selected using systematic sampling based on a sampling interval determined from the sampling fraction and a random start (provided by the CDC). The patient list was considered wrap-around, so that if a selected patient's medical records could not be found, the next patient on the list at the sampling interval was chosen.

Sample size. The total sample size was chosen for each study site based on statistical considerations (precisions of estimates of interest) and available resources at the local level. Each study site targeted a total sample size of approximately 300 patients. This sample size would allow calculation of an estimate of a proportion with a 95% confidence interval (CI) ± 9.8%, assuming a design effect of 3.0.

The sample of 300 patients was allocated among the provider size strata, guided by the proportions of all living patients from providers of the various sizes obtained from surveillance data. Approximately half the sample for the large providers was allocated to providers receiving RWCA support and half to providers not receiving RWCA support.

Data collection. For selected patients, medical records at the sampled facilities were reviewed to determine if the patient was in care during 1998, as well as to determine information about prescribed medications, diagnosed illnesses, laboratory test results, and service utilization (inpatient, outpatient, and emergency room visits). If information was available in the primary medical record about care received at other providers (e.g., copies of records from a hospital admission or a consultation report), those data were collected and treated as care received during the period. If the same individual was selected at multiple providers, information about care was abstracted only from the medical records of the provider at which the individual was first selected. This study did not attempt to collect information from all facilities at which the individual accessed care during 1998, due to the large numbers of providers potentially involved and resource constraints.

Human subjects. The study was determined by the CDC to be non-research. Of the three participating state and local health departments, the protocol was reviewed and received Institutional Review Board (IRB) approval in two, and in one, it was determined to be exempt from IRB review.

Analysis

Calculation of sampling weights. The data analysis used methods that accounted for the stratified two-stage cluster design and the variation in the selection probabilities for patients from different providers. Because selection of patients within a provider was based on a systematic sample with a random start, all patients selected from the same provider and same race/sex stratum had equal selection probabilities. To account for these selection probabilities, the weight for each patient was the inverse of the probability that the individual was sampled. This probability was computed as the probability of sampling the provider at which the patient received care, multiplied by the probability of sampling the patient at the provider.

Calculation of weighted point estimates, standard errors, and confidence intervals. Using SAS PROC SURVEY MEANS, we calculated point estimates (means, proportions), their standard errors (SEs), and 95% CIs.[12] For some measures (such as number of outpatient visits in the study interval), median values were estimated. Estimates of medians and their SEs, and 95% CIs were calculated using PROC DESCRIPT in SUDAAN with DESIGN=WR (with replacement).[13]

As a result of the cluster sample design, the variance of an estimate was usually larger than the variance from a simple random sample of the same size. The magnitude of the increased variance due to clustering was measured as the *design effect,* defined as the ratio of variance incorporating the sampling design to the variance from a simple random sample of the same size.[14] We calculated the design effect for several health care outcome measures of interest.

Comparison of SHDC data with data from other surveillance activities

Adult and Adolescent Spectrum of HIV Disease Project. The ASD is a multicenter medical record review of approximately 100 clinics in 10 U.S. cities; its methods have been described.[6] A standardized instrument was used to collect information from medical records on the patients' clinical conditions and treatments at baseline and every six months until the patient's death or until the last contact with the patient at one of the participating facilities. In this analysis, we included data abstracted in the Seattle, Detroit, and New Orleans sites for care received during 1998.

Certain clinical and demographic characteristics have been shown previously to be associated with the extent to which patients are treated according to treatment guidelines.[15,16] Since demographic characteristics of patients could differ for ASD and SHDC, estimates for ASD health care measures were standardized to the SHDC population of patients receiving medical care in the Detroit, New Orleans, and Seattle metropolitan areas. Estimates were standardized according to the weighted proportion of patients in SHDC by race (black, Hispanic, "other"), HIV risk mode and sex (men who have sex with men [MSM], injecting drug user [IDU] male, IDU female, other risks male, other risks female), and age group (<30 years, ≥30 years).

HIV/AIDS Reporting System data. In order to compare the demographic characteristics of individuals observed in the SHDC project to the characteristics of individuals diagnosed with HIV infection (with or without AIDS) in the same project areas, we used local HIV/AIDS surveillance data. HARS is implemented in all U.S. states, with well-characterized completeness of reporting.[17] We used local surveillance data reported through January 2003 to calculate numbers of individuals with HIV diagnoses known to be living with HIV at any point in 1998 (diagnosed before December 31, 1998, and not known to be dead as of December 31, 1997).

Definitions of standards of care

In order to describe the extent to which care for HIV infection complied with standards of practice, we used United States Public Health Service/Infectious Disease Society of America (USPHS/IDSA) treatment guidelines in effect during 1998.[18] Based on these guidelines, specific definitions for standards of care to be evaluated in our study were established and appear on this page (see right).

Definitions of standards of care

Based on the United States Public Health Service/Infectious Disease Society of America Guidelines for the Prevention of Opportunistic Infections in Individuals Infected with HIV, the following definitions were used during the Survey of HIV Disease and Care Project (SHDC) to evaluate the extent to which care complied with standards of practice.

Highly active antiretroviral therapy (HAART)

Met standard: Prescription of AZT+(ddI, ddC, or 3TC)+(any protease inhibitor or any non-nucleoside reverse transcriptase inhibitor), D4T+(ddI or 3TC)+ (any protease inhibitor or any non-nucleoside reverse transcriptase inhibitor), combivir+(any protease inhibitor or any non-nucleoside reverse transcriptase inhibitor), abacavir+AZT+ddC, abacavir+ddI+3TC, or trizivir+(ddC or ddI or any protease inhibitor or any non-nucleoside reverse transcriptase inhibitor). Protease inhibitors included saquinavir, ritonavir, indinavir, nelfinavir, amprenavir, tipranavir, kaletra. Non-nucleoside reverse transcriptase inhibitors included delavirdine, nevirapine, efavirenz, emivirine, capravirine.

Eligibility: History of an AIDS-defining opportunistic illness diagnosis, CD4<500, PCR>20000, or bDNA>10,000

Pneumocystis carinii pneumonia (PCP) prophylaxis

Met standard: Prescription of dapsone, pentamidine, or trimethoprim-sulfamethoxazole

Eligibility: History of PCP diagnosis or CD4<200

Mycobacterium avium (MAC) prophylaxis

Met standard: Prescription of clarithromycin, azithromycin, or rifabutin.

Eligibility: History of M. avium diagnosis or CD4<50 cells/μL

Tuberculosis skin test (TST)

Met standard: Any tuberculin skin test during 1998

Eligibility: No tuberculosis diagnosis or sputum culture positive to M. tuberculosis prior to 1998

Influenza vaccine

Met Standard: Received influenza vaccination during 1998

Eligibility: All individuals observed

RESULTS

Of 51 providers approached to participate in the project, four (8%) declined participation. The sites were not consistent in their refusal rates; King County had 75% participation (12 of 16 providers participated), and the other two sites had 100% participation. Medical records were reviewed for 915 HIV-infected patients receiving medical care at participating facilities in 1998. The number of records abstracted per site ranged from 253 to 374 patients per site.

Demographic characteristics of the 915 patients observed in SHDC and of the 5,266 patients followed during 1998 by the ASD Project in the same states are shown in Table 1; these data should be interpreted with the understanding that in Michigan and Louisiana, the ASD projects are conducted in Detroit and New Orleans, but the SHDC data represent the entire state of Michigan and the southern health districts of Louisiana. The modal age group for subjects in both projects was 30–39 years in all areas except the ASD project in Michigan, where the modal age group was 40–49 years; most subjects in all areas were male. The proportions of the SHDC sample who did not have a major HIV risk factor documented in their medical records ranged from 25% to 50%; the proportions of patients observed in ASD who did not have this information ranged from 9% to 20%.

The distribution of demographic characteristics of individuals reported to HARS in the study areas of King County, health districts 1,2,3,4 and 9 in Louisiana, and Michigan are compared with the weighted SHDC sample in Table 2. For all project areas, higher proportions of women were observed by SHDC (11% to 31%) than by HARS (8% to 26%). In all three project areas, the most common age group of

Table 1. Demographic characteristics of patients observed in SHDC[a] and patients enrolled in ASD[b] in southern Louisiana, Michigan state, and King County, Washington, during 1998

	King County		*Louisiana*		*Michigan*	
	SHDC (N=288)	*ASD (N=1,211)*	*SHDC (N=253)*	*ASD (N=2,702)*	*SHDC (N=374)*	*ASD (N=1,412)*
Characteristic	n *(percent)*	n *(percent)*	n *(percent)*	n *(percent)*	n *(percent)*	n *(percent)*
Sex						
Male	178 (62)	991 (82)	158 (62)	1,926 (71)	258 (69)	903 (64)
Female	110 (38)	220 (18)	95 (38)	776 (29)	116 (31)	509 (36)
Age						
13–19	2 (1)	9 (1)	16 (6)	39 (1)	9 (2)	14 (1)
20–29	51 (18)	160 (13)	43 (17)	473 (18)	40 (11)	153 (11)
30–39	132 (46)	590 (49)	111 (44)	1,194 (44)	151 (40)	486 (34)
40–49	74 (26)	357 (29)	58 (23)	764 (28)	123 (33)	565 (40)
50+	29 (10)	95 (8)	25 (10)	232 (9)	51 (14)	194 (14)
Race						
White, non-Hispanic	143 (50)	738 (61)	92 (36)	946 (35)	175 (47)	355 (25)
Black, non-Hispanic	84 (29)	248 (20)	152 (60)	1,670 (62)	179 (48)	1,010 (72)
Hispanic	30 (10)	144 (12)	7 (3)	78 (3)	9 (2)	37 (3)
Asian/Pacific Islander	17 (6)	31 (3)	1 (0)	8 (0)	0 (0)	5 (0)
American Indian/Alaska Native	11 (4)	50 (4)	0 (0)	0 (0)	2 (1)	0 (0)
Other/unknown	3 (1)	0 (0)	1 (0)	0 (0)	9 (2)	5 (0)
Risk category						
MSM	105 (36)	667 (55)	61 (24)	999 (37)	94 (25)	607 (43)
MSM/IDU	18 (6)	179 (15)	13 (5)	262 (10)	6 (2)	87 (6)
IDU	27 (9)	150 (12)	46 (18)	435 (16)	52 (14)	365 (26)
HRH	67 (23)	99 (8)	42 (17)	428 (16)	35 (9)	229 (16)
Other/unknown	71 (25)	116 (10)	91 (36)	531 (20)	187 (50)	124 (9)
Clinical/immunological						
CD4 0–99 or AIDS-defining opportunistic illness	108 (38)	424 (35)	87 (34)	1,177 (44)	167 (45)	684 (48)
CD4 100–349	78 (27)	444 (37)	56 (22)	823 (30)	58 (16)	447 (32)
CD4≥350 or unknown	102 (35)	343 (28)	110 (43)	702 (26)	149 (40)	281 (20)

NOTE: The columns for geographic areas are not directly comparable because for Louisiana and Michigan, the SHDC and the ASD collected data in different (but overlapping) geographic areas. In Michigan, the geographic area of the cohort study was limited to the Detroit metropolitan area, but the geographic area of the sampling study was the entire state. In Louisiana, the geographic area of the cohort study was limited to the New Orleans metropolitan area, but the geographic area of the sampling study was several health districts in the southern third of the state.

[a]The Survey of HIV Disease and Care Project (SHDC) is a probability sample of individuals in care for HIV infection.

[b]The Adult and Adolescent Spectrum of HIV-related Disease Study (ASD) is a facility-based cohort study of individuals with HIV infection.

MSM = men who have sex with men

IDU = injection drug user

HRH = high risk heterosexual

individuals observed in SHDC or reported to HARS was 30–39 years. In Michigan, individuals observed in SHDC were older than individuals reported to HARS (48% vs. 29% older than 39 years). The proportions of patients in King County and Michigan with other or no identified transmission risk factors for HIV in the SHDC data were much higher than those reported to HARS in those same project areas (24% vs. 7% for King County; 46% vs. 9% for Michigan).

An evaluation of the provision of care in accordance with guidelines in effect in 1998 is presented in Table 3. In most cases, the point estimates for care parameters from ASD sites were within the 95% CIs determined in SHDC, with some exceptions. For example, in King County, the SHDC estimates of proportions of individuals with at least one documented viral load assay or CD4 measurement were higher than proportions from individuals observed in ASD. The opposite was true in Michigan.

Similarly, health care services utilization data indicated that the point estimates for utilization parameters in ASD were in most cases within the 95% CIs calculated by SHDC methods (see Table 4). ASD patients in King County had more viral load testing and less CD4 testing per six months

Table 2. Demographic characteristics of patients observed in SHDC[a] and patients with HIV infection reported to the local HIV/AIDS Reporting System (HARS) in southern Louisiana, Michigan state, and King County, Washington, during 1998

	King County		*Louisiana*		*Michigan*	
	SHDC (N=288)	*HARS (N=4,121)*	*SHDC (N=253)*	*HARS (N=8,633)*	*SHDC (N=374)*	*HARS (N=9,532)*
Characteristic	*Weighted percent*	*percent*	*Weighted percent*	*percent*	*Weighted percent*	*percent*
Sex						
Male	89	92	69	74	70	77
Female	11	8	31	26	30	23
Age						
13 –19	0	0	4	0	2	1
20–29	13	9	15	19	11	29
30–39	40	45	45	41	40	41
40–49	34	33	26	29	34	22
50+	13	12	10	10	14	7
Race						
White, non-Hispanic	77	76	35	34	32	35
Black, non-Hispanic	13	13	61	63	61	59
Hispanic	5	7	4	2	4	4
Asian/Pacific Islander	3	2	1	0	0	0
American Indian/Alaska Native	1	2	0	0	0	0
Other/unknown	0	0	0	0	2	1
Risk category						
MSM	53	72	29	32	26	45
MSM/IDU	9	7	5	7	4	5
IDU	6	10	28	19	15	16
HRH	8	4	12	13	9	14
Other/unknown	24	7	26	27	46	19
Clinical/immunological						
CD4 0–99 or AIDS-defining opportunistic illness	39	NC	30	NC	43	NC
CD4 100–349	25	NC	32	NC	14	NC
CD4≥350 or unknown	36	NC	38	NC	43	NC

NOTE: The columns for geographic areas are directly comparable. For both the SHDC and HARS, data were collected from patients in the same geographic areas. In Michigan, the entire state was included; in Louisiana, health districts 1,2,3,4, and 9 were included; and in Washington, King County was included.

[a]The Survey of HIV Disease and Care Project (SHDC) is a probability sample of individuals in care for HIV infection.

NC = not collected for many HARS cases; therefore data are incomplete and not reported

MSM = men who have sex with men

IDU = intravenous drug user

HRH = high-risk heterosexual

than patients observed in SHDC, while ASD patients in Michigan had more CD4 testing per six months than SHDC patients. The median number of outpatient visits as estimated by SHDC ranged from four to six per year, with no significant differences between ASD and SHDC.

The design effects for some outcomes of interest are presented in Table 5. Considerable variation occurred in design effects across outcomes and across states. Most notably, the design effect for influenza vaccination in Michigan was very high, and average cluster sizes in Michigan were higher than in the other project areas. Observed design effects of <1.0 indicated that the sampling approach resulted in smaller variance than would be expected with a simple random sample of the same size.

DISCUSSION

The SHDC pilot study demonstrated the feasibility of a two-stage sampling approach to obtain population-based estimates of indicators of quality of care for HIV-infected patients. The collection of clinical surveillance information from the medical records of individuals in care for HIV

Table 3. Proportion of patients treated according to guidelines among patients observed in SHDC[a] and patients enrolled in ASD[b] in southern Louisiana, Michigan state, and King County, Washington, during 1998

	Patients receiving treatment					
	King County		*Louisiana*		*Michigan*	
	SHDC (N=288)	*ASD (N=1,211)*	*SHDC (N=253)*	*ASD (N=2,702)*	*SHDC ((N=374)*	*ASD (N=1,412)*
Care provided	*Treated/eligible[c] (Weighted percent [95% CI])*	*Treated/eligible[c] (Standardized percent)*	*Treated/eligible[c] (Weighted percent [95% CI])*	*Treated/eligible[c] (Standardized percent)*	*Treated/eligible[c] (Weighted percent [95% CI])*	*Treated/eligible[c] (Standardized percent)*
HAART[d]	166/242 (68 [57, 79])	751/1,119 (64)	111/171 (64 [58, 70])	1,508/2,457 (59)	161/249 (69 [54, 85])	949/1,308 (75)
PCP prophylaxis[e]	122/149 (79 [60, 99])	436/544 (83)	77/108 (73 [54, 92])	1,130/1,402 (80)	144/174 (86 [81, 90])	696/843 (83)
MAC prophylaxis[f]	28/50 (64 [46, 82])	132/204 (67)	20/47 (57 [42, 71])	465/647 (70)	37/77 (61 [42, 80])	257/457 (61)
TB test[g]	164/287 (55 [37, 74])	871/1,147 (70)	111/253 (62 [31, 92])	1,144/2,278 (50)	129/373 (34 [10, 58])	488/829 (52)
Influenza vaccine[h]	82/288 (27 [17, 37])	311/1,211 (24)	65/253 (27 [19, 35])	686/2,702 (26)	50/374 (21 [0, 45])	383/1,412 (28)
Viral load assay[h]	264/288 (93 [89, 97])	1,013/1,211 (83)	156/253 (73 [47, 99])	2,019/2,702 (74)	164/374 (47 [27, 67]	1,156/1,412 (84)
CD4 measurement[h]	271/288 (95 [91, 100])	1,024/1,211 (82)	175/253 (78 [55, 100])	2,223/2,702 (82)	238/374 (68 [50, 86])	1,197/1,412 (87)

NOTE: The columns for geographic areas are not directly comparable because for Louisiana and Michigan, the SHDC and the ASD collected data in different (but overlapping) geographic areas. In Michigan, the geographic area of the cohort study was limited to the Detroit metropolitan area, but the geographic area of the sampling study was the entire state. In Louisiana, the geographic area of the cohort study was limited to the New Orleans metropolitan area, but the geographic area of the sampling study was several health districts in the southern third of the state.

[a]The Survey of HIV Disease and Care Project (SHDC) is a probability sample of individuals in care for HIV infection.

[b]The Adult and Adolescent Spectrum of HIV-related Disease Study (ASD) is a facility-based cohort study of individuals with HIV infection.

[c]Treated/eligible is the number of individuals receiving care/number of individuals eligible for care.

[d]Eligibility for highly active antiretroviral therapy (HAART) was defined as CD4<500, PCR>20,000, bDNA>10,000, or AIDS-opportunistic illnesses diagnosis (any time before or during 1998).

[e]Eligibility for Pneumocystis carinii pneumonia (PCP) prophylaxis was defined as CD4<200, PCP diagnosis (any time before or during 1998).

[f]Eligibility for Mycobacterium avium complex (MAC) prophylaxis was defined as CD4<50 or Mycobacterium avium diagnosis (any time before or during 1998).

[g]Eligibility for tuberculosis (TB) test was defined as no TB diagnosis or sputum culture positive to M. tuberculosis prior to 1998.

[h]All individuals are recommended to receive an annual influenza vaccine and regular viral load assay and CD4 T-lymphocyte (CD4) measurements.

CI = confidence interval

infection is critical to the evaluation of provision of appropriate care. Previously, the CDC-supported ASD project was used to describe trends in HIV-related morbidity and natural history of HIV infection, as well as to inform national surveillance case definitions and treatment guidelines.[7,8,19–24] The SHDC method represents a model for collection of data for similar purposes with increased representativeness.

Our study differs in two important ways from HCSUS, which surveyed a large national probability sample of individuals in care for HIV infection.[4,5,25] First, HCSUS was generalizable to the entire United States because it randomly sampled providers from throughout the United States, while SHDC was not because it selected three geographic areas through competitive application. Second, the sampling frame of providers for SHDC benefited by using the HIV/AIDS surveillance database as a sampling frame. The national reporting system for HIV/AIDS is regularly evaluated. Since AIDS is reportable to all state and local health departments and AIDS case subject reporting is known to have high completeness, investigators in SHDC had a very complete sampling frame of providers of care for individuals with HIV infection.[26]

Table 4. Characteristics of health care utilization of patients observed in SHDC[a] and patients enrolled in ASD[b] in southern Louisiana, Michigan state, and King County, Washington, during 1998

	Patients receiving treatment											
	King County				*Louisiana*				*Michigan*			
	SHDC (N=288)		*ASD (N=1,217)*		*SHDC (N=253)*		*ASD (N=2,711)*		*SHDC (N=374)*		*ASD (N=1,356)*	
Utilization measure	*Median*	*Weighted median (95% CI)*	*Median*	*Standardized median*	*Median*	*Weighted median (95% CI)*	*Median*	*Standardized median*	*Median*	*Weighted median (95% CI)*	*Median*	*Standardized median*
Outpatient visits[c]	6	5.9 (5.2, 7.2)	6.9	5.9	4	4.0 (3.3, 6.2)	3.9	3.8	3	4.0 (1.6, 6.3)	4.6	4.6
Viral loads[d]	1.5	1.5 (1.4, 1.5)	1.6	1.7	0.7	1.0 (0.2, 1.4)	1.0	0.9	0.5	0.4 (0, 1.4)	1.5	1.4
CD4 counts[d]	1.5	1.5 (1.4, 1.5)	1.1	1.1	1	1.0 (0.5, 1.4)	1.0	1.0	0.5	0.6 (0, 1.1)	1.1	1.3

NOTE: The columns for geographic areas are not directly comparable because for Louisiana and Michigan, the SHDC and the ASD collected data in different (but overlapping) geographic areas. In Michigan, the geographic area of the cohort study was limited to the Detroit metropolitan area, but the geographic area of the sampling study was the entire state. In Louisiana, the geographic area of the cohort study was limited to the New Orleans metropolitan area, but the geographic area of the sampling study was several health districts in the southern third of the state.

[a]The Survey of HIV Disease and Care Project (SHDC) is a probability sample of individuals in care for HIV infection.

[b]The Adult and Adolescent Spectrum of HIV-related Disease Study (ASD) is a facility-based cohort study of individuals with HIV infection.

[c]Number of visits in the study year

[d]Number of tests per six months

The advantage of our design is that, in theory, the results should provide population-based estimates that represent all individuals in care for HIV infection in the participating geographic areas. However, some providers selected from the sampling frame did not allow review of their patients' medical records. This problem was encountered in King County, where 75% of selected providers participated. In the HCSUS study, 70% to 100% of selected providers participated. The extent to which such refusals bias the results of the study depends on whether the providers who refused to participate (*1*) serve a substantial proportion of HIV-infected patients in care, (*2*) serve substantively different patient populations, or (*3*) provide a substantively different quality of care. In our study, most refusals were providers who had few HIV-infected patients.

The SHDC study design has additional limitations. The design only allows observation of and inference to individuals in care for HIV infection; however, individuals diagnosed with HIV infection but not in care are a critical population.[27] With respect to analysis of eligibility for standard care measures, some criteria could not be determined with confidence because we did not access the full medical history to determine previous diagnoses or procedures. Finally, because we did not abstract records from all care providers for each patient, our data represent minimum estimates of care received during the year.

Because we sampled providers with probability proportion to size and without replacement, the variance estimates could have been obtained by taking into account joint selection probabilities and selection without replacement (e.g., using SUDAAN with design=UNEQWOR).[13] However, we chose to analyze data as though the sampling had been done with replacement by using SAS PROC SURVEY MEANS and SUDAAN with design=WR (with replacement), as previously done in other studies.[12,13,28,29] This analytic approach results in somewhat larger variance estimates; however, previous comparisons using pilot data (data not shown) demonstrated the resulting increases in variance to be a small and acceptable trade-off for the use of simplified analytic methods and the ability to use SAS software, which is widely available to state health departments.[12]

Design effects measure the extent to which the sampling design increased the variance of the estimates compared with using simple random sampling; larger design effects result in wider CIs. The variation in design effects that we observed in our study was considerable, and we will conduct additional investigation into the reasons for these variations as we prepare for subsequent studies. It is possible that there were variations among the participating study sites in the extent to which sampled providers represented single practitioners (which would tend to result in higher intra-class correlation and higher design effects) vs. groups of physicians (which would tend to result in more heterogeneity of practice patterns and lower intra-class correlation). Future study designs will increase the number of sampled providers and reduce the number of sampled patients per provider in order to reduce design effects.

Future cycles also may include an interview component, which would provide the opportunity to evaluate psychosocial factors not included in the medical records. Offering

Table 5. Intra-class correlation coefficients, design effects, and related design characteristics from SHDC[a] in southern Louisiana, Michigan state, and King County, Washington, during 1998

Region Variable	*Number of observations*	*Mean cluster size*	*Number of clusters*	*Coefficient of variation for weights*	*Intra-class correlation coefficient (ρ)[b]*	*Design effect for proportion or mean[c]*
King County (*n*=288; *m*=140; *ddf*=134)						
HAART[d] (%)	242	2.02	120	1.087	0.1558	1.87
PCP prophylaxis[e] (%)	149	1.99	75	1.054	0.2741	6.16
Influenza vaccine[f] (%)	288	2.06	140	1.106	0.0282	2.19
Viral loads per six months[g] (mean)	288	2.06	140	1.106	0.0870	1.41
Louisiana (*n*=253; *m*=68; *ddf*=63)						
HAART[d] (%)	171	3.17	54	0.973	0.0043	0.87
PCP prophylaxis[e] (%)	108	2.77	39	0.950	−0.0329	0.52
Influenza vaccine[f] (%)	253	3.72	68	1.012	0.1065	1.80
Viral loads per six months[g] (mean)	253	3.72	68	1.012	0.0551	2.25
Michigan (*n*=374; *m*=21; *ddf*=13)						
HAART[d] (%)	249	12.4	20	1.161	0.1018	5.76
PCP prophylaxis[e] (%)	174	9.2	19	1.190	−0.0531	0.57
Influenza vaccine[f] (%)	374	17.8	21	1.153	0.4062	29.62
Viral loads per six months[g] (mean)	374	17.8	21	1.153	−0.0431	0.81

[a]The Survey of HIV Disease and Care Project (SHDC) is a probability sample of individuals in care for HIV infection.

[b]ρ estimates were obtained using SUDAAN Proc Regress with R=Exchangeable and model variable=sex.

[c]Design effects were obtained using SUDAAN with option DEFT1 and DESIGN=WR.

[d]Eligibility for highly active antiretroviral therapy (HAART) was defined as CD4<500, PCR>20,000, bDNA>10,000, or AIDS-opportunistic illnesses diagnosis (any time before or during 1998).

[e]Eligibility for Pneumocystis carinii pneumonia (PCP) prophylaxis was defined as CD4<200, PCP diagnosis (any time before or during 1998).

[f]All individuals are recommended to receive an annual influenza vaccine and regular viral load assay and CD4 T-lymphocyte (CD4) measurements.

[g]SUDAAN cannot calculate design effect for medians.

n = total number of charts abstracted

m = number of primary sampling units

ddf = denominator degrees of freedom

interviews to individuals with no known care provider might help identify barriers to entry into care and determine needs for health services among individuals not in care for HIV infection.

We conclude that conducting two-stage probability sampling in the setting of health department HIV surveillance programs is possible. Our estimates of some indicators of HIV care were comparable with those of the currently conducted facility-based clinical surveillance project, but for important outcomes such as intensity of laboratory monitoring of CD4 counts, differences were observed. Health departments are uniquely positioned to conduct such surveillance projects because only they have well-validated sampling frames of providers of HIV care and because only they have statutory authority to conduct medical record reviews for individuals with HIV infection.

Beginning in 2006, CDC expects to support a random sample of states to collect clinical outcomes surveillance data by a method similar to SHDC. This will represent a three-stage cluster sampling design that can extend inference to all patients in care for HIV infection in the United States. Further evaluation of these methods in more health departments should help to clarify the feasibility of using this study design in diverse public health settings. Larger numbers of patients must be included to provide locally useful data. To the extent that this methodology is feasible, adaptable, and cost effective, it may provide an important clinical surveillance tool for descriptive epidemiology and program evaluation, both nationally and by states and major metropolitan areas.

Acknowledgements are due to the following authors who comprise the SHDC Pilot Study Group whose input contributed to this study: Hien L. Guyer, MPH, and Linda L. Wotring, PhD, of HIV/AIDS Surveillance, Michigan Department of Community Health, Detroit and Lansing, MI; A.D. McNaghten, PhD, Paul Jones, MS, Debra L. Hanson, MS, John T. Brooks, MD, David Swerdlow, MD, Mitchell I. Wolfe, MD, MPH, and Maxine Denniston, MS, of the Division of HIV/AIDS Prevention-Surveillance and Epidemiology, National Center for HIV, STD, and TB Prevention, Centers for Disease Control and Prevention,

Atlanta, GA. The authors thank Erik Schwab for developmental editing and copyediting of the manuscript and Amy Drake for assistance with project management.

REFERENCES

1. Lee LM, Karon JM, Selik R, Neal JJ, Fleming PL. Survival after AIDS diagnosis in adolescents and adults during the treatment era, United States, 1984–1997. JAMA 2001;285:1308-15.
2. Report of the NIH panel to define principles of therapy of HIV infection. MMWR Recomm Rep 1998;47(RR-13):1-41.
3. Summers T, Alagiri P, Kates J. Federal HIV/AIDS spending: a budget chartbook, fiscal year 2002. Menlo Park (CA): The Henry J. Kaiser Family Foundation; 2003. Also available from: URL: http://www.kff.org/hivaids/loader.cfm?url=/commonspot/security/getfile.cfm&PageID=21717 [cited 2005 Jan 20].
4. Shapiro MF, Berk ML, Berry SH, Emmons CA, Athey LA, Hsia DC, et al. National probability samples in studies of low-prevalence diseases. Part I: Perspectives and lessons from the HIV cost and services utilization study. Health Serv Res 1999;34(5 Part 1):951-68.
5. Frankel MR, Shapiro MF, Duan N, Morton SC, Berry SH, Brown JA, et al. National probability samples in studies of low-prevalence diseases. Part II: designing and implementing the HIV cost and services utilization study sample. Health Serv Res 1999;34(5 Part 1):969-92.
6. Farizo KM, Buehler JW, Chamberland ME, Whyte BM, Froelicher ES, Hopkins SG, et al. Spectrum of disease in individuals with human immunodeficiency virus infection in the United States. JAMA 1992;267:1798-805.
7. Jones JL, Hanson DL, Dworkin MS, Alderton DL, Fleming PL, Kaplan JE, et al. Surveillance for AIDS-defining opportunistic illnesses, 1992–1997. MMWR CDC Surveill Summ 1999;48(2):1-22.
8. McNaghten AD, Hanson DL, Jones DL, Dworkin MS, Ward JW. Effects of antiretroviral therapy and opportunistic illness primary chemoprophylaxis on survival after AIDS diagnosis. Adult/Adolescent Spectrum of Disease Group. AIDS 1999;13:1687-95.
9. Louisiana Department of Health and Hospitals. Administrative regions [cited 2005 Jan 15]. Available from: URL: http://www.dhh.louisiana.gov/offices/medialibrary/media-1/REG_MAP04.jpg
10. Department of Health and Human Services (US), HIV/AIDS Bureau. Ryan White CARE Act [cited 2005 Jan 15]. Available from: URL: http://hab/hrsa.gov/history.htm
11. Insightful Corporation. S-PLUS: Version 6. Washington: Insightful Corporation; 2003.
12. SAS Institute. SAS: Version 8.2. Cary (NC): SAS Institute; 2002.
13. Research Triangle Institute. SUDAAN user manual, release 8.0. Research Triangle Park (NC): Research Triangle Institute; 2001.
14. Cochran WG. Sampling techniques. New York: John Wiley & Sons; 1977.
15. Moore RD, Stanton D, Gopalan R, Chiasson RE. Racial differences in the use of drug therapy for HIV disease in an urban community. N Engl J Med 1994;330:763-8.
16. Hsu LC, Vittinghoff E, Katz MH, Schwarcz SK. Predictors of use of highly active antiretroviral therapy (HAART) among individuals with AIDS in San Francisco, 1996–1999. J Acquir Immune Defic Syndr 2001;28:345-50.
17. Klevens RM, Fleming PL, Li J, Gaines CG, Gallagher K, Schwarcz K, et al. The completeness, validity, and timeliness of AIDS surveillance data. Ann Epidemiol 2001;11:443-9.
18. United States Public Health Service and Infectious Disease Society of America. USPHS/IDSA guidelines for the prevention of opportunistic infections in individuals infected with human immunodeficiency virus [cited 2005 Jan 15]. Available from: URL: http://aidsinfo.nih.gov/guidelines/default_db2.asp?id=69
19. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR Recomm Rep 1992;41 (RR-17):1-19.
20. Kaplan JE, Hanson DL, Jones JL. Early initiation of combination antiretroviral therapy: does it affect clinical outcome? Proceeding of the 13th International AIDS Conference; 2000 July; Durban, South Africa. Abstract No.: LbPeB7051.
21. Dworkin MS, Hanson DL, Kaplan JE, Jones JL, Ward JW. Risk for preventable opportunistic infections in individuals with AIDS after antiretroviral therapy increases CD4+ T lymphocyte counts above prophylaxis thresholds. J Infect Dis 2000;182:611-5.
22. Kaplan JE, Hanson JL, Navin TR, Jones JL. Risk factors for primary Pneumocystis carinii pneumonia in human immunodeficiency virus-infected adolescents and adults in the United States: reassessment of indications for chemoprophylaxis. J Infect Dis 1998;178:1126-32.
23. Brooks JT. Risk of Mycobacterium avium complex (MAC) infection among HIV-infected individuals who stop primary prophylaxis after HAART-associated immune reconstitution. Proceedings of the Infectious Diseases Society of America 40th Annual Meeting; 2000; San Diego (CA). Abstract No.: 500.
24. Blair J, Hanson DL, Jones JL. Do gender differences in viral load predict differences in HIV disease progression? Proceedings of the 7th Conference on Retroviruses and Opportunistic Infections; 2000 Feb; San Francisco. Abstract No.: 195.
25. Bozzette SA, Berry SH, Duan N, Frankel MR, Leibowitz AA, Lefkowitz D, et al. The care of HIV-infected adults in the United States. HIV Cost and Services Utilization Study Consortium. N Engl J Med 1998;339:1897-904.
26. Nakashima AK, Fleming PL. HIV/AIDS surveillance in the United States, 1981–2001. J Acquir Immune Defic Syndr 2003;32 Suppl 1:S68-S85.
27. Centers for Disease Control and Prevention. HIV prevention strategic plan through 2005 [cited 2005 Jan 15]. Available from: URL: www.cdc.gov/nchstp/od/hiv_plan
28. Hansen MH, Hurwitz WN, Madow WG. Sample survey methods and theory. New York: John Wiley & Sons; 1953.
29. Kish L. Survey sampling. New York: John Wiley; 1965.