Part B: Statistical Methods

No statistical methods are used to complete the Annual Performance Report. The discussion below refers to statistical methods for the Annual Homeless Assessment Report.

B1   Potential Respondent Universe for the Annual Homeless Assessment Report

Per the HMIS Data Standards, program providers in each CoC collect data on program participants at participant entry and exit and, for some data elements, at least once annually during program enrollment, if the period between program entry and exit exceeds one year.

Data are typically collected manually and then entered into the local HMIS by staff of the recipient organization. However, recipients may also collect and enter data into HMIS simultaneously or "in real time." To complete a local Annual Homeless Assessment Report, recipients must aggregate data contained in HMIS into AHAR response tables associated with each question. In many instances this is accomplished automatically through a canned report in HMIS.

HUD expects and encourages all CoCs (approximately 448) to submit local AHAR reports. The burden estimates in section A.12 are based on this assumption. However, when HUD initiated this effort in 2002, some CoCs did not have a functional HMIS and many others were collecting information from only a portion of homeless assistance providers in their community. As a result, HUD developed a representative national sample of 102 communities to target for data collection until all CoCs could participate in the AHAR. Participation has steadily grown beyond the original sample. For the 2008 report, 222 CoCs contributed their data.

B2   Statistical Methods and Data Collection Procedures

B.2.1   Sampling Plan

This section describes the procedures for selecting a nationally representative sample of 102 jurisdictions for the AHAR.[1]

---

[1]   The initial AHAR sample consisted of 80 jurisdictions. Some jurisdictions from the original sample—especially jurisdictions representing rural areas—were unable to provide data to the AHAR because of HMIS implementation issues or other data quality concerns. In addition, several of the rural sample sites did not have any homeless residential service providers located in their jurisdiction. As a result, we were unable to report data by geography. In an effort to improve the scope and quality of data from rural jurisdictions, 22 additional rural jurisdictions were added to the AHAR sample. Thus, there are a

CDBG Jurisdictions Are the Primary Sampling Units

The AHAR uses the geographic areas defined for the allocation of CDBG funds as the primary sampling unit. The four types of CDBG jurisdictions are:

- Principal cities[2]

- Cities with 50,000 or more persons (that are not principal cities)

- Urban counties

- Rural areas or non-entitlement jurisdictions

CDBG jurisdictions constitute the basic building blocks of CoCs. In some cases, the CDBG jurisdiction and the CoC represent the same geographic area (e.g., principal cities are often a single CoC), but, in other situations, the CDBG jurisdiction is a geographic subunit of the CoC (e.g., a small city with 50,000 or more persons may be a subunit of a countywide CoC). The selection of 102 CDBG jurisdictions ensures the inclusion of a wide range of sites in the AHAR as well as the reasonably precise measurement of the characteristics of homeless persons and their patterns of service use.

HUD provided a sampling frame for the selection of CDBG jurisdictions. The sampling frame is a list of all 3,142 CDBG jurisdictions within the 430 CoCs in the 50 states as of 2002.[3] The next section describes the decision to stratify the sites based on geographic type, along with the procedures for selecting certainty and non-certainty sites.

---

total of 102 AHAR sample sites.

[2] The original file from which the sample was selected used the category of "central city" for CDBG jurisdictions rather than "principal city." However, the CDBG program moved to designation of principal city rather than central city following the OMB guidance, and the definition of central city and principal city are slightly different (see 24 CFR Part 570). Of the 482 CDBG central city jurisdictions that existed both before and after the definition change, 327 central city jurisdictions (68%) became principle cities with the definition change. A small number of non-central cities (85 out of 2,501) in the original file were categorized as principal cities in the 2007 CDBG file. In our analysis by CDBG jurisdiction and in procedures for adjusting the sampling weights, we used the community's current CDBG jurisdiction to ensure that our results accurately represented the current system for designating CDBG jurisdictions.

[3] HUD provided a file called "COC_GeoAreasInfo.xls" with a list of 3,219 CDBG jurisdictions, jurisdiction type, and population of each jurisdiction. Geographic areas in the U.S Territories and Puerto Rico and three duplicate records were eliminated, resulting in a sampling frame of 3,142 CDBG jurisdictions. In addition, four CDBG areas in Massachusetts and one in New Hampshire included overlapping geographic areas and double-counted the population; therefore, the population was evenly divided across the overlapping CDBG jurisdictions before sampling.

Stratifying the Sample by Type of Geographic Area

A CDBG jurisdiction may be a large principal city of a metropolitan area, a smaller city with a population of 50,000 or more, one or more suburban or urban fringe counties, or a rural area. As such, the number of homeless persons in each jurisdiction varies considerably.

Using the relative size of the homeless population in each CDBG jurisdiction to select a sample may increase the precision of the estimates for any particular sample size. However, with the number of homeless persons in each CDBG jurisdiction unknown, the study team assumed that the total population in each CDBG jurisdiction provided a measure of relative size of the homeless population for purposes of sample selection. The study team premised the assumption on the likelihood that the number of homeless persons is correlated with the total population in the area served by the CDBG jurisdiction. The team further refined the assumption by dividing the sample into strata based on the expected rate of homelessness.[4]

Earlier research on homelessness indicates that the rate of homelessness varies by type of geographic area. For example, Burt (2001) found that 71 percent of the homeless persons using homeless-related services are located in principal cities but that only 30 percent of the total U.S. population lives in principal cities.[5] By contrast, rural areas account for 9 percent of the homeless population, but 20 percent of the overall population. Further, suburban/urban fringe areas represent 21 percent of homeless persons, but 50 percent of the overall population. These findings suggest that, before using the total population as a proxy for the relative size of the homeless population, the CDBG jurisdictions should be stratified by type of geographic area to account for the fact that the ratio of homeless persons to the population varies across geographic areas. Hence, the study team divided the CDBG jurisdictions into four groups based on their classification for the allocation of CDBG funds: principal cities, other cities larger than 50,000, urban counties, and rural areas (i.e., counties that are part of non-entitlement areas). Such stratification increases the precision of estimates.

---

[4]   Sampling based on the expected rate of homelessness is an attempt to obtain more precise estimates than those yielded by a simple random sample. If the proxy for the expected rate of homelessness is not correlated with the actual rate of homelessness, the resulting estimates will still be unbiased; however, the extra precision gains go unrealized.

[5]   Burt, Martha. 2001. Homeless Families, Singles, and Others: Findings from the 1996 National Survey of Homeless Assistance Providers and Clients. *Housing Policy Debate*, V12 (4), 737-780. This report presents the share of the homeless population by urban/rural status. The share of the population in each type of geographic area comes from the author's calculations based on March 1996 Current Population Survey data. The results from the Burt study were based on central cities rather than principal cities, but we refer to them as principal cities here because of the high degree of overlap and to make the discussion easier to follow.

Very Large CDBG Jurisdictions Selected with Certainty

Given that the size of the population across CDBG jurisdictions is skewed by a few very large jurisdictions covering areas with several million residents, a useful strategy for reducing sampling variability in the estimated number and characteristics of homeless persons is to select very large jurisdictions in the sample with certainty. Selecting a CDBG jurisdiction with certainty means that the CDBG jurisdiction represents only itself in the sample estimates but ensures that the sample does not exclude the largest jurisdictions, whose number and characteristics of the homeless population could substantially affect national estimates. Exhibit B-1 lists the 18 CDBG jurisdictions selected with certainty.

For selecting the certainty sites, the study team divided the CDBG jurisdictions into the four geographic-type strata. Assuming the rate of homelessness was the same in each area within each stratum, the study team calculated the standard deviation (square root of the variance) of the number of homeless persons for the entire stratum. The team then recalculated the standard deviation by excluding the largest site (as if that site were taken with certainty) to obtain a relative estimate of the reduction in the variance of the estimates that would occur if that site were selected with certainty. In the event of substantial reduction in the variance due to the selection of the certainty unit, the overall variance of the sample estimates will be smaller as the variance contribution to the estimate from the certainty sites is zero. The process of selecting the next-largest site as a certainty site continued until the reduction of the variance or standard deviation was small or marginal. The process resulted in the identification of 11 certainty sites consisting of eight principal cities, one other city larger than 50,000, and two urban counties (but no non-entitlement areas).

Based on earlier research findings showing that homeless persons are disproportionately located in principal cities, the study team identified 7 additional principal cities as certainty sites, for a total of 15 principal cities in the certainty sample (and 18 certainty sites in total). The team selected the 7 additional principal cities with certainty because the cities had among the largest populations of persons living in emergency and transitional shelters in the 1990 and 2000 Census counts.[6] All 7 certainty sites had one of the 10 largest counts in either 1990 or 2000.[7] Given that so many homeless persons live in these cities, it is important to include them with certainty in a nationally representative sample.

---

[6]   For 1990 counts, see U.S. Department of Housing and Urban Development. "Allocating Homeless Assistance by Formula." A Report to Congress, 1992. For 2000 counts, see U.S. Census Bureau. "Emergency and Transitional Shelter Population: 2000." A Census 2000 Special Report.

[7]   The other 8 certainty sites in principal cities were all ranked in the top 15 in the 1990 or 2000 Census counts.

	Geographic Area	Type of CDBG Entity	Size of Housed Population	Census Region	CoC Name
		Exhibit B-1: Geographic Characteristics and Population of 18 Certainty Sites			
1	NEW YORK CITY	Principal City	8,008,278	Northeast	New York City Coalition/CoC
2	LOS ANGELES	Principal City	3,694,820	West	County of Los Angeles, CA
3	CHICAGO	Principal City	2,896,016	Midwest	Chicago CoC
4	HOUSTON	Principal City	1,953,631	South	Houston/Harris County
5	PHILADELPHIA	Principal City	1,517,550	Northeast	City of Philadelphia
6	PHOENIX	Principal City	1,321,045	West	Maricopa CoC