2022 Annual Homelessness Assessment Report to Congress:

PART 2 METHODOLOGY REPORT

DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT
JUNE 2024

Contents

Introduction

This document summarizes the methodology for producing the 2022 Annual Homeless Assessment Report (AHAR). Specifically, it describes the data sources, how we process the data, and how we weight to estimate the national count for all CoCs when only a subset of CoCs provided usable data in the Longitudinal Systems Analysis Report (LSA).

This report has three sections: First, we summarize how we clean the data and determine whether the CoC's data are usable; the second section describes how we combine and prepare the LSA and HIC data for analysis; and the third section describes the steps for calculating the weights by project-household category so that the data from usable CoCs in each category can be weighted to account for all people served in all CoCs.

Data Sources

The 2022 AHAR is based on two primary sources of data:

1. *Homeless Management Information Systems (HMIS).* The HMIS data for 2022 cover a one-year reporting period, October 1, 2021, to September 30, 2022 as captured in the LSA. The data contain information on people who used emergency shelters, safe havens, transitional housing, rapid re-housing, and permanent supportive housing at any point during this period. HMIS data are unduplicated at the project-level and reported in the aggregate for each project.
2. *Housing Inventory Count (HIC).* HIC data represent the number of beds and units available for people experiencing homelessness on a single night. For several CoCs, inventory data in HMIS collected through the LSA upload had data quality issues. In these cases, project-level HIC data was used to help estimate the number of beds available for occupancy.

Data Cleaning and Determination of Data Usability

There were three primary phases of data cleaning that occurred before the production of the national estimates – the upload phase, the data review phase, and the usability phase. These are described below.

LSA Report Upload Phase

Communities uploaded the 10 csv files comprising their LSA report directly into the HDX 2.0. Beginning in 2020, two types of upload error-checks were programmed directly into the data collection portal, resulting in automatic file rejections for specific issues. If a file was rejected, communities worked with their HMIS vendor to resolve issues and upload a new LSA report.

The first set of error-checks was for "fatal" upload errors, or those that represent major violations of the LSA Programming Specifications. The second set of error-checks was for upload errors that were not due to issues with the programming specifications used to produce the LSA file. These error checks flagged contradictions in the data. Some examples include scenarios where one field in the data indicates a client was not served in a specific project type, but another field classifies them as a head of household in that project type, or the minimum age reported in a specific project type is greater than the maximum age reported for the same project type. Since data to create the LSA reports were extracted directly from each CoC's HMIS, any errors that fell within this second set of error-checks were corrected directly in each CoC's HMIS by the CoCs. Some communities also worked with their vendors to resolve these types of issues.

After the CoC fixed the upload errors, they uploaded a new (corrected) LSA report to the HDX 2.0. Once the LSA report was successfully uploaded, the CoC moved into the data review phase.

Data Review Phase

The data quality review phase began following the successful upload and submission of an LSA report in the HDX 2.0. In 2022, this review examined tens of thousands of potential data quality issues across CoCs' LSA uploads. Data quality issues were divided into two categories: data errors and data warnings. Errors were impossibilities in the data. Warnings identified issues that were technically possible but were either unlikely or that may represent data quality issues.

Both error and warning checks are programmed directly into the HDX 2.0 and run immediately after the LSA files were successfully uploaded. This provided communities with a near-instant overview of their upload's entire data quality. To assist communities in resolving these issues, errors and warnings were divided into two categories: those that likely required vendor support (e.g., due to issues related to the program developed to produce the LSA report) and those that likely reflected issues in HMIS. Communities worked with an assigned Data Liaison to understand and respond to any issues that were identified, as well as with their vendors if resolving the issues required their support.

Data cleaning was an iterative process, often involving multiple LSA uploads and data review files. CoCs were given a deadline to respond to any data quality issues flagged by their Data Liaison. Changes to data cannot be made manually within HDX or in the csv files included in the upload itself, so resolving errors or warnings in the LSA required CoCs to either: 1) explain why the data were accurate as-is by providing notes for warnings in the HDX, or 2) fix any issues within their HMIS and re-upload a new LSA report to the HDX 2.0.

Usability Determination Phase

Once data are final, each CoC's data were assessed for whether they could be used in the report and could be used to impute for missing data (either at the project or CoC level).[1] This step is referred to as "usability determination." Data were reviewed for usability for nine categories defined by both the project type (Emergency Shelter/Transitional Housing/Safe Havens (EST); Rapid Rehousing (RRH); and Permanent Supportive Housing (PSH)) and the household type (Adult Only (AO); Adult and Child (AC); and Child Only (CO)), (see Exhibit 1).

Senior members of the Data Review Team reviewed data quality issues at both the CoC level (the nine usability categories) and the individual project level to determine if the data submitted was of high enough quality to be included in the AHAR. The data for the CoC could be determined usable at the project-household category if the data submitted for projects within that category were absent major data quality issues.

The Data Review Team looked for many indicators of potential data quality issues, included high numbers of overlapping enrollments (which would lead to duplication), considerable issues related to reported inventory (affecting utilization rates), projects with large numbers of clients missing an enrollment location, projects with large numbers of households with either more than one assigned head of household or less than one head of household, CoCs missing entire projects, unusually high or low utilization rates (the proportion of a homeless provider's beds occupied by clients, on average and at several points in time), and unexplained and unusually long or short program stays. To use as much data

[1] Missing data that needs to be imputed for includes both data for projects that do not participate in HMIS and data for projects in CoCs for which data quality was an issue.

as possible, usability determinations were expanded to include fully usable, people data usable, and not usable. These determinations were made at the project type and household type level. For example, if a CoC had data quality issues with their EST data for adult-only households, EST data on families was still usable if those data did not have the same data quality issues.

Many of the data quality issues in 2022 were related to inventory data. Where data on people served were of high enough quality (i.e., in the people data usable category) but inventory data were not, HIC data were used in place of LSA inventory data to estimate utilization (discussed in more detail in section Constructing the Bed Count below).

Table 1 Number of CoCs with Usable Data by Project and Household Type, 2022

	EST-AO	EST-AC	EST-CO	RRH – AO	RRH - AC	RRH - CO	PSH – AO	PSH – AC	PSH - CO
Fully Usable	198	201	200	258	257	260	212	211	228
People Data Usable	49	45	43	14	11	12	41	40	28
Not Usable	139	140	143	114	118	114	133	135	130
Total CoCs	386	386	386	386	386	386	386	386	386

Source: 2022 Longitudinal System Analysis.
Note: "Fully Usable" in the context of Usability Determination means that after the usability assessment of a CoC's submitted HMIS data, all projects and all people within that reporting category were deemed acceptable to use in the report.

In 2022, 322 CoCs (or 83 percent of CoCs) had usable data in at least one project and household type category and 64 CoCs (or 17 percent of CoCs) had no usable data. This reflects an improvement over 2021, for which 72 percent had any usable data and 28 percent had no usable data.

Compiling and Preparing Data for Estimation Procedures

After the usability determination phase and before constructing the weights, data are compiled and prepared for estimation procedures in two phases. The phases are outlined below, followed by a summary of each phase:

1. A data preparation phase that:
    a. Reads in data sources, including the raw LSA data, usability determination information, and the HIC
    b. Corrects inconsistencies in project information as well as person- and household-level data
    c. Implements CoC-level corrections to usable bed data
    d. Assigns household types to unclassified households using both person- and household-level data
2. A merge phase that:
    a. Creates a project crosswalk by matching LSA projects to HIC projects
    b. Merges HIC data to LSA data using the crosswalk
    c. Adjusts the HIC for any LSA Splits (i.e., where single HIC project corresponds to multiple LSA projects)
    d. Compiles four separate datasets based on the original data source to be referenced in later estimation procedures

i. Data Source 1: HIC-only
ii. Data Source 2: LSA-only
iii. Data Source 3: Merged LSA-HIC Data
iv. All Sources Compiled: Includes Data Sources 1-3 grouped by CoC and reporting category

In the Preparation Phase, data was imported in the following order: raw LSA data, usability determination information, and HIC data. At each import step, basic checks were performed (e.g. confirm data were imported correctly and data expected to be there are there) to assess LSA upload information across CoCs, to determine frequency counts of usability categories, and to derive counts of HIC beds based on project and household type combinations.

The next step involved cleaning adjustments for LSA Person data, LSA Household data, and project IDs. Corrections targeted systematic inconsistencies found in the data with household type, with age and parenting variables, with the demographic universe, and with household project type statuses. Other adjustments made in this step of the preparation phase were corrections to errors in the data. For example, we made manual fixes to geography type when the project information did not match official HUD guidance. Also, for the 2022 cycle, several CoCs submitted data with Project UUIDs instead of Project IDs. With CoCs' assistance, we created a crosswalk of UUIDs to project IDs.

The preparation phase then assigns unclassified households. An unclassified household means the household type cannot be determined based on the data given for members of that household. Age may be missing for a household member for example, which makes it difficult to tell if the household is Adult Only or an Adult-Child household. Unclassified household assignment is done by using LSA Person data to distinguish between unclassified households never served in a classified household versus unclassified households served in at least one classified household.

Unclassified households that were served in at least one classified household are instances where there is some information known about the household type. They may have been served in one or multiple project groups where the household type was classified. This known household type information is scanned to determine whether it is appropriate to assign the same household type to the unclassified household based on the available information from other project groups.

Unclassified households that were never served in a classified household have no information about their household type. They may have been served in one or multiple project groups, but in each of those groups the household type was unclassified. For each CoC/project type/race/ethnicity combination, we create household type assignment targets based on household type likelihoods for that combination. Then, within the CoC/project type/race/ethnicity classes, we randomly select unclassified households to classify as Adult Only, Adult and Child, and Child Only based on the predefined assignment targets.

Out of 670,878 total households, 9,484 unclassified households were randomly assigned a household type. In other words, approximately 1.4 percent of households underwent this household type assignment process. The resulting household type assignments were then saved in an updated version of the LSA Household dataset.

In the **Merge Phase**, we created a crosswalk between the LSA and the HIC that resulted in a common HIC ID for each LSA project—a crucial piece necessary for downstream analysis. This was done by matching LSA projects to HIC projects using direct matching from both current and previous years' data and results from fuzzy matching from previous years' data.

With the crosswalk completed, we could then process the HIC data to merge it to the LSA data. Again, HIC data were critical in CoCs where LSA inventory data was not of high enough quality to determine utilization rates (people per bed), a necessary part of estimate production. An additional step adjusts HIC data based on "LSA Splits"—instances where a single HIC project corresponds to multiple LSA projects. In this step, we determined how HIC beds counts should be distributed across the LSA projects. The preference was for beds to be distributed according to the split share of average people counts. If the average and total people count were zero for an "LSA Split" HIC project, then the beds were distributed uniformly across splits. This adjusted data was then recombined with non-split observations. As a final step before weight construction, data were identified as coming from three distinct sources: HIC Only, LSA Only, and Merged LSA-HIC.

A Note on Reclassification

During the merge phase, projects with discrepancies in people counts, HMIS beds, and non-HMIS beds were identified and their bed counts reclassified. Specifically, we looked for instances where projects had data on people, but there were no HMIS beds for the project and non-HMIS beds were non-zero. This discrepancy indicated that the project was likely misclassified as "HMIS non-participating" and required reclassification.

In FY2022, LSA reports did not include data on people from projects with only non-HMIS beds, and thus no projects in the LSA-only data source required reclassification. Within the Merged LSA-HIC data source, if a project had LSA data on people accessing projects and had only non-HMIS beds on the HIC, the project's HIC beds were reclassified. In this reclassification, HMIS beds were set to equal the non-HMIS bed number, and the non-HMIS bed number is then set to zero, essentially setting the project to HMIS-participating. These reclassifications were made under the assumption that HMIS beds were mis-reported as non-HMIS by CoCs on the HIC.

In 2022, 36 projects in the Merged LSA-HIC data source were reclassified from non-HMIS to HMIS-participating, representing less than 1 percent of all projects.

Constructing the Weights

Summary

Once data were prepared, weights were constructed for the participating projects so that the data for participating projects also represented people who stayed in non-participating projects. Participating projects are (a) in CoCs that were marked usable in the usability determination phase and (b) reported HMIS participating beds (including those reclassified to having HMIS beds). Non-participating projects are either (a) in CoCs marked as not usable in the usability determination phase or (b) reported no HMIS participating beds. The following summarizes the steps below for constructing the weights.

The number of people served (or people counts) are the ultimate units for calculating the weights. For people counts in participating projects, we use the number of people reported using the project's beds. For people counts in non-participating projects, the number of people was estimated by imputing persons per bed counts based on known data from participating projects of the same category.[2] To tie people counts in non-participating projects to observed bed data, we imputed people per bed rather than people only. Imputed people per bed is then multiplied by observed beds to arrive at the estimated number of people in

[2] The characteristics of the category varied by match criteria as described below in the section on imputation of persons per bed at non-participating projects.

non-participating projects. The total number of people across participating and non-participating projects becomes the numerator for the weight. The actual number of people in participating projects is the denominator for the weight.

When estimating aggregated person counts across participating and non-participating projects in the CoC, people who use more than one project within a CoC will be counted more than once. To correct for this overcounting, we calculate the percentage of people overlapping in projects in a fully participating CoC and estimate this percentage for CoCs not fully participating. We then adjust the CoC-level person count by this project-overlap percentage.

After adjusting for overlap across projects within a CoC, we compare project-level and CoC-level LSA datasets and correct for any inconsistencies. Once inconsistencies between these LSA datasets are corrected, we identify CoCs in which people are in more than one household type during the reporting year. People can be in more than one household type either by joining or leaving a household with children or aging into adulthood.[3] After identifying household overlap, we construct weighting classes within each project type and household type (including people in multiple household types). After weighting classes are constructed, the weights can be computed. The weight is defined as the estimated total count of people divided by the reported count of people by weighting class within each project-household group. Reported person counts, drawn from fully participating projects, will be applied to weights so that the total count is an estimate for all CoCs whether participating or not.

The following sections detail the steps for calculating the weights, including imputing missing users per bed, estimating project and household overlap, constructing weight classes, and computing and applying the final weights.

Constructing the Bed Count

Bed counts are used for estimating people from non-participating projects. The people reported in participating projects and the estimate of people using non-participating projects are the units for calculating weights. Therefore, bed counts are required from both participating and non-participating projects to estimate the number of people using non-participating projects. Bed counts were obtained from a combination data from the Longitudinal System Analysis (LSA) and the Housing Inventory Counts (HIC).

Person and bed inventory counts were drawn from the LSA database if they were determined to be usable. If a CoC's bed inventory counts from the LSA were determined not to be usable in any specific project-household category, inventory data were instead taken from the HIC.

Table 2 displays the number of projects and beds from each data source by project-household type for all projects in the AHAR universe—fully participating projects, partially participating projects (projects in which some, but not all, beds are HMIS-participating), and fully non-participating projects. Table 2 also lists the percentage of beds from each data source. There are indications that LSA data quality is improving. Compared to 2021, a larger percentage of projects, across all project groups and household types, contributed both usable person and bed inventory information in the LSA while a smaller share contributed neither of these two types of data.[4] Table 3 similarly displays the number of projects, beds, and percentage of beds from each data source by project-household group but for just the projects that

[3] If a person in a child only household turns 18 during a reporting year, they are now considered part of an adult only family if they are not a parent or guardian. If the youngest child in a household turns 18 during the reporting period, they are now part of an Adult Only household rather than a household composed of Adults and Children.
[4] Equivalent tables displaying project counts from the 2021 AHAR can be found in the AHAR 2021 Methodology Report.

were either fully or partially participating.[5] Compared to 2021, a larger share of projects provided usable bed inventory counts in the LSA among projects that provided usable person counts.

Table 2. Number of Projects-Household Group Beds by Data Source for all Projects

Project Type	Household Type	HIC Only			LSA Only			Both HIC and LSA		
		# of Projects	# of Beds	% of All Beds	# of Projects	# of Beds	% of All Beds	# of Projects	# of Beds	% of All Beds
EST	AO	3,621	105,584	42	3,907	118,602	46	1,058	29,283	12
	AC	2,001	63,965	44	2,070	80,144	45	529	14,470	12
	CO	205	1,461	41	248	1,406	50	48	389	10
PSH	AO	2,670	102,451	36	3,898	132,663	53	791	24,229	11
	AC	1,455	57,279	42	1,592	51,577	46	418	12,646	12
	CO	4	31	100	0	0	0	0	0	0
RRH	AO	1,487	28,962	32	3,111	40,355	68	0	0	0
	AC	1,142	34,224	31	2,600	59,177	70	0	0	0
	CO	16	52	39	25	49	61	0	0	0

Source: 2022 Housing Inventory Counts and 2022 Longitudinal System Analysis. Note: EST includes Emergency Shelters, Safe Havens, and Transitional Housing projects; PSH are Permanent Supportive Housing projects; and RRH are Rapid Re-housing projects. % of all beds is the percent of beds in that project and household type category.

Table 3. Number of Project-Household Group Beds by Data Source for Participating Projects

Project Type	Household Type	LSA Only			Both HIC and LSA		
		# of Projects	# of Beds	% of All Beds	# of Projects	# of Beds	% of All Beds
EST	AO	3,177	97,318	80	796	22,815	20
	AC	1,753	69,638	83	373	11,191	18
	CO	211	1,140	84	41	344	16
PSH	AO	3,262	100,042	85	581	17,382	15
	AC	1,456	47,116	83	305	9,239	17
	CO	0	0	0	0	0	0
RRH	AO	3,066	39,733	100	0	0	0
	AC	2,553	58,229	100	0	0	0
	CO	25	49	100	0	0	0

Source: 2022 Housing Inventory Counts and 2022 Longitudinal System Analysis. Note: EST includes Emergency Shelters, Safe Havens, and Transitional Housing projects; PSH are Permanent Supportive Housing projects; and RRH are Rapid Re-housing projects. % of all beds is the percent of beds in that project-type and household type.

Imputation of Persons per Bed at Non-Participating Projects

The weights are defined as the estimated unduplicated count of persons at all CoCs in the weighting class in the reporting category divided by the reported number of unduplicated persons for all CoCs in the

[5] Partially participating projects are included in the project count, but only the participating portion of beds in these projects are included in the number and percentage of beds.

weighting class in the reporting category.[6] To estimate the unduplicated counts of persons at all CoCs, we first need to impute persons per bed in non-participating projects.

We employed a stratified mean-based imputation scheme for persons per bed. This involved establishing the following hierarchy of project classifications:

1. Project Type
2. Household Type
3. Project Geography (Rural/Suburban/Urban)
4. Beds above Median (i.e., whether the project has more beds than the median bed count for projects with the same project type, household type, and geography type).

When imputing persons per bed, we looked to projects that shared the above classifications (here, we refer to "cells," i.e., all projects which share identical values for the classifications listed above). If there were at least 30 projects with available persons per bed data in a given cell,[7] we used the mean persons per bed value amongst projects with non-missing data in the cell to impute for those projects with missing data in the cell. If there were less than 30 projects with non-missing persons per bed in a cell, then we relaxed our matching criterion, removing the requirement that all projects in the cell share the same "Beds above Median" status. We continued to relax the matching criterion until there were at least 30 projects with non-missing data in the cell (where we relaxed according to the hierarchy above, removing the "least important" conditions first).

Table 4 below displays the frequency of the "match level" of projects for which we needed to impute persons per bed; i.e., how often we had to resort to various levels in the hierarchy. A project with a match level 1 with a missing users per bed value used the mean of projects with matching project type, household type, geography, and beds above median status to impute for this value. Almost all matches (97.7 percent) were match level 1, indicating that all four match variables were used for the imputation. As can be seen in the table, 1.1 percent used only one match variable—project type— for the imputation (match level 4).

Table 4. Distribution of Match Level for Projects with Missing Persons per Bed

Match Level	Characteristics Matched On	# of Projects	% of Projects
1	project type household type geography beds relative to median status	14,942	97.7
2	project type household type geography	122	0.8
3	project type household type	67	0.4
4	project type	166	1.1

Source: 2022 Housing Inventory Counts and 2022 Longitudinal System Analysis.
Note: There are 15,297 projects missing persons per bed.

[6] Weight classes group projects within each project-household type with similar characteristics together so that non-participating projects are represented by participating projects with similar characteristics. Details on how CoCs were grouped into weighting classes can be found in the section below on Constructing Weighting Classes under Finalizing Weights.

[7] For the purposes of imputation, we removed outliers within each cell, where outliers have persons per bed more than three standard deviations below or above the mean for the cell in question. After removing outliers, we required that a cell have at least 30 projects with non-missing data to impute using that cell.

New York City (NY-600) was handled separately from the other CoCs given its uniqueness in size. Therefore, we carried out this process separately for the NYC projects (with the only changes being switching project type and household type in the hierarchical order and reducing the minimum cell size from 30 to 10). In short, we allowed New York City to represent only itself.

Table 5 displays the distribution of reported persons per bed among participating projects by project and household type while table 6 displays the distribution of imputed persons per bed among non-participating projects. In general, this imputation methodology tightened the overall (participating and non-participating projects) distribution of persons per bed relative to the distribution of participating projects, increasing the 25th percentile and decreasing the 75th percentile. This is to be expected, as using a mean-based imputation approach ensures that the same value will be used to impute missing persons per bed data at many projects. The analysis team explored using methods that would preserve the underlying distribution of reported persons per bed; however, missing data rates and the idiosyncratic nature of the data rendered such techniques highly volatile.

Table 5. Distribution of Reported Persons per bed Among Participating Projects by Project and Household Type

Project Type	Household Type	# of Projects	Mean	Minimum	25th Percentile	50th Percentile	75th Percentile	Maximum
ES	AO	2,786	6.3	0.0[a]	2	3.6	6.6	338.5
	AC	1,619	5.4	0.0	1.6	2.7	5	590.5
	CO	205	6.3	0.0	1	3.7	9	51
TH/SH	AO	1,187	2	0.0	1.1	1.7	2.4	28
	AC	507	1.6	0.0	0.9	1.4	2.1	10.8
	CO	47	1.6	0.0	0.7	1.3	2	6.5
PSH	AO	3,843	1.1	0.0	0.9	1	1.2	37
	AC	1,761	1	0.0	0.7	1	1.2	11
	CO	0	0	0.0	0	0	0	0
RRH	AO	3,066	2.7	0.0	0.9	1.5	3	71
	AC	2,553	2.5	0.0	0.9	1.4	2.6	165
	CO	25	0.7	0.0	0	0	0	6

Source: 2022 Housing Inventory Counts and 2022 Longitudinal System Analysis.

Note: ES are Emergency Shelters, TH/SH are Transitional Housing/Safe Havens projects; PSH are Permanent Supportive Housing projects; and RRH are Rapid Re-housing projects. AO is Adults Only; AC is Adults and Children; and CO is Children Only. There are 17,599 participating projects across all project and household types.

[a] A value of 0.0 users per bed indicates the project in that project-household type did not serve any people in the reporting year.

Table 6. Distribution of Imputed Persons per bed Among Non-Participating Projects by Project and Household Type

Project Type	Household Type	# of Projects	Mean	Minimum	25th Percentile	50th Percentile	75th Percentile	Maximum
ES	AO	3,101	5.2	3.3	4.8	4.8	5.1	9.7
	AC	1,753	4.3	2.2	3.8	4.7	4.7	6.5
	CO	182	5.7	2.8	5.1	5.8	6.9	6.9
TH/SH	AO	1,512	1.9	1.7	1.8	1.9	1.9	2.1
	AC	721	1.5	1.2	1.5	1.5	1.6	1.9
	CO	67	1.5	1.5	1.5	1.5	1.5	1.9
PSH	AO	3,516	1.1	0.9	1	1.1	1.1	1.3
	AC	1,704	1	0.8	0.9	1	1.1	1.1
	CO	4	3.5	1	2.6	4.3	4.3	4.3
RRH	AO	1,532	2.3	1.4	1.4	3	3	3.5
	AC	1,189	2.2	1.3	1.3	2.6	3.1	3.1
	CO	16	4.3	4.3	4.3	4.3	4.3	4.3

Source: 2022 Housing Inventory Counts and 2022 Longitudinal System Analysis.
Note: ES are Emergency Shelters, TH/SH are Transitional Housing/Safe Havens projects; PSH are Permanent Supportive Housing projects; and RRH are Rapid Re-housing projects. AO is Adults Only; AC is Adults and Children; and CO is Children Only. There are 15,297 non-participating projects across all project and household types.

After estimating persons per bed in non-participating projects, the estimated number of persons in non-participating projects can be derived. The estimated count of persons in non-participating projects is defined as the product of the observed number of beds and the estimated count of persons per bed.

Estimating Percent Overlap across Participating and Non-Participating Projects at the CoC Level within Reporting Categories

Persons utilizing more than one project within a reporting category in a CoC will be counted more than once when aggregated across projects.[8] To correct for this over-counting, an estimate of the percentage overlap across projects in a CoC is calculated. When the estimates are aggregated across projects in a reporting category, we reduce the sum by this overlap percentage.

For calculating this overlap percentage within reporting categories, the observed overlap percentage is first calculated across participating projects for each fully participating and partially participating CoC.[9] The overlap percentage among participating projects is defined as the total number of duplicated persons minus the total number of unduplicated persons divided by the total number of duplicated persons in a fully or partially participating CoC.

The overlap percentages among CoCs with participating projects is then used to estimate overlap percentages among CoCs which are either non-participating or partially participating.[10] To do so, we employed a modified predictive mean matching scheme to impute overlap percentages from CoCs with a

[8] There are nine reporting categories, defined by each combination of project type (Emergency Shelter/Safe Haven/Transitional Housing, Rapid Re-Housing, and Permanent Supportive Housing) and household type (Adults Only, Adults and Children, and Children Only).
[9] All projects are participating in a fully participating CoC. Some, but not all, projects are participating in a partially participating CoC.
[10] For CoCs fully participating in a given year, the actual observed overlap rate was used, so no imputation needed.

participation rate of 50 percent or more.[11] This imputation scheme followed the steps below for imputing project-level overlap for non-participating and partially participating CoCs:

1. To calculate the unduplicated number of people in a fully participating CoC, we first calculate the number of people overlapping projects in a participating CoC.[12] To do this, we multiply the reported duplicated count of people by the calculated percentage of duplicated people in overlapping projects. The calculated number of people overlapping projects is then subtracted from the reported duplicated count of people served in the CoC.
2. Use of previous year's data to impute for current year's data – In the event that a CoC had available overlap rates in the previous year, but not the current year (i.e., less than 50 percent of the CoC's projects were participating in that reporting category in current year), we used the previous year's data to impute for the current year.
3. For CoCs in which less than 50 percent of projects were participating but overlap rates from the previous year were not available, we used predictive mean matching to construct an estimate for the overlap rate. The data from the model were based on other CoCs that were available from the current year. If the current year was not available from the other CoCs, the overlap rates from the previous year were used if available. Predictive mean matching is a hotdeck imputation procedure that involves the following steps:
    a. Construct a linear regression model using a specified set of predictors to predict overlap rate (where the model is built on cases with non-missing overlap rate).[13]
    b. Estimate the predicted overlap rate (based on the model created above) for all CoCs.
    c. For a given CoC with a missing overlap rate, take the five CoCs (with reported overlap rates) that have the most similar predicted overlap rate to the given CoC. Then randomly assign one of their reported overlap rates to the given CoC as its imputed value.
    d. Repeat the previous step for all CoCs with missing overlap rates.
4. For CoCs with more than 50 percent of projects participating we use the reported duplication rate as long as the reported duplication rate is greater than the total (estimated) duplication rate.[14]
5. To estimate the unduplicated number of people in a non-participating CoC, we multiply the estimated number of duplicated people by the estimated overlap rate. This is the estimated number of overlapping people. The product is then subtracted from the estimated number of duplicated people to arrive at the unduplicated number of people in a non-participating CoC. In a partially participating CoC, we first estimate the duplicated count of people served in the partially participating CoC.[15] This number is then subtracted by either the calculated number of people overlapping across participating projects in the partially participating CoC or the estimated number of people overlapping across all projects in the CoC, whichever is greater.

[11] The participation rate equals the number participating projects in a CoC divided by the total number of projects in a CoC.

[12] People overlapping projects are people in more than one project in the same CoC within the reporting year.

[13] Partially participating CoCs are used to fit the estimation model and included among the CoCs being estimated. However, only the participating portion of partially participating CoCs is utilized for fitting the estimation model. When imputing for partially participating CoCs, the imputation is for the entirety of the CoC (both the participating and non-participating portions).

[14] If the total estimated duplication rate is greater, we take the estimated rate, with one exception: it's possible the higher estimated duplication rate will lead to the total count being less than participating. If this occurs, we make an adjustment. We subtract the number of duplicated users in participating projects from the estimated total number of duplicated users and multiply this difference by the estimated total overlap rate. This becomes the estimated number of overlapping users in non-participating projects. We then add the estimated number of overlapping users in non-participating projects to the calculated number of overlapping users in participating projects. This sum is then divided by the estimated number of duplicated users in all projects to arrive at the total estimated duplication rate.

For partially participating CoCs, the duplicated count of people served is the sum of estimated duplicated people in non-participating projects and reported duplicated people in participating projects.

Addressing Inconsistencies across Project and CoC-level Data Sources

The components for calculating the weights—person counts in participating projects and person counts in all (participating and non-participating) projects—are constructed from project level data. In addition to project level data, we have reported CoC-level person counts. In some situations, project-level person counts aggregated to the CoC-level (the counts that make up the weight components) will suggest no participation (i.e., zero people experiencing homelessness) while another LSA dataset at the CoC-level will suggest participation (i.e., greater than zero people experiencing homelessness).[16] In these situations, the CoC's participation status and person counts used for calculating the weights are adjusted from non-participating to participating and the person counts are adjusted to a value greater than zero. To calculate this adjustment factor, we first divide the participating person count aggregated from the project level across all CoCs by the participating person count aggregated from the CoC level across all CoCs. This national adjustment factor is then multiplied by the CoC level participating person count at the CoCs with this inconsistency, which then becomes the participating person component at these CoCs for estimating the weights. The additional participating person component is included in the weight calculations.

Estimating Household-Type Overlap at the CoC-Level

When aggregating persons across household types, failure to adjust for household-type overlap leads to counting the same household more than once if the household was in more than one household type in the reporting year.

There are three household types in the 2022 AHAR: households that are adult only (AO), households that have adult(s) and children (AC), and households that only have children (CO). CoCs can be composed of projects with one or more of these household types. For example, a CoC may have AO projects only or it may have AO, AC, and CO projects. Household-type overlap can take place if a household was one household type (e.g., Adults Only) at a point during the reporting year and a different household type (e.g., Adults and Children) later in the reporting year. Since we independently estimate the number of people experiencing homelessness by household type, when we add across household-type categories to get the total number of people experiencing homelessness, the people in multiple household types during the year are double counted or duplicated. Here we describe our procedures for estimating this overlap to ensure a person who is in more than one household type during the year is only counted once when we add estimates across household types.

In the end, this will allow us to estimate the number of unduplicated people in participating and non-participating projects in each of the seven possible household-type combinations (AO Only, AC Only, CO Only, AO and AC Only, AO and CO Only, AC and CO Only, and AO, AC, and CO). These estimates can be added up to estimate the number of unduplicated people experiencing homelessness in the CoC or experiencing homelessness in any single or combined category of household types. The rest of this section describes how we calculate the number of unduplicated people in these seven household-type combinations for participating projects (part A) and then how we estimate the numbers in these household-type combinations for participating and non-participating projects together (part B). Both sets of numbers are needed for weighting our data.

Part A. Calculate unduplicated count of household-type overlap for people in participating projects by CoC.

[16] One percent (RRH-AC) to twelve percent (EST-CO) of project-household types indicate this inconsistency.

From our data on participating projects in a CoC, we know the duplicated count of people in each of the seven household type combinations and the percentage of all duplicated people in each of these household type combinations.[17] We use this information to calculate the unduplicated number of people who are in a single household type or overlap multiple household-type categories.

1. **Calculate the number of people that overlap participating AO, AC, and CO household-type projects**. To do this, we first calculate the duplicated number of people in participating projects who are in AO, AC, and CO household types. To calculate this duplicated count, multiply (a) the duplicated number of people in participating projects in any of the three household type categories by (b) the percentage of duplicated people in participating projects who are in AO, AC, and CO projects.[18] We divide this number by 3 to arrive at the unduplicated number of people in participating projects that are in all three household types.[19]

2. **Calculate the number of people that only overlap participating AO and AC household-type projects**. If the CoC has AO and AC projects but no CO projects, the overlap is calculated by multiplying (2) the number of duplicated people in AO or AC projects by (b) the percentage of duplicated people in AO or AC projects that are in both AO and AC projects. We divide this number by 2 to arrive at the estimated unduplicated number of people that are in both AO and AC projects. If the CoC has AO, AC, and CO projects, the number of people overlapping AO and AC projects is determined by first calculating the duplicated number of people in multiple household types. To do this, we multiply (a) the total number of duplicated people across all three household types by (b) the percentage of duplicated people in any household type who are in multiple household types. We then subtract the duplicated count of people in AO, AC, and CO projects (the overlap count calculated in step 1 prior to dividing by 3). This duplicated count of people served in multiple household types, excluding people in all three household types, is then multiplied by the percentage of duplicated people in multiple household types (also excluding people in all three household types) who are in AO and AC projects. As with CoCs with no CO projects, we divide by 2 to arrive at the estimated unduplicated number of people in both AO and AC projects.

3. **Calculate the number of people that only overlap participating AO and CO household-type projects**. If the CoC has AO and CO projects but no AC projects, the overlap is calculated by multiplying (a) the number of duplicated people in AO or CO projects by (b) the percentage of duplicated people in AO or CO projects that are in both AO and CO projects. We divide this number by 2 to arrive at the estimated unduplicated number of people that are in both AO and CO projects. If the CoC has AO, AC, and CO projects, the number of people overlapping AO and CO projects is determined by first calculating the duplicated number of people in either (a) AO and

[17] The duplicated counts and percentages for the seven household type combinations are based on discrete CoC level data. These percentages are then applied to the weight components which were aggregated from the project level to the CoC level.
[18] The percentage of duplicated people in participating projects who are in AO, AC, and CO projects is equal to the duplicated number of observed people in AO, AC, and CO household types divided by the total number of duplicated people across all seven household combinations.
The weight components are constructed from the project level. Household overlap counts are not available at the project level, but they are available at the CoC level. Therefore, overlap percentages are calculated at the CoC level and then applied to aggregated data originally from the project-level.
[19] The calculated overlap between household types cannot be larger than the number of people in any of the overlapping household types. If the overlap is larger, the minimum of these household types is used instead. For example, if the calculated overlap between AO, AC, or CO household types is larger than the number of people in AO, AC, or CO households, the overlap is set to the minimum value of AO, AC, or CO households.

CO household types or in (b) AC and CO household types. To do this, we subtract the duplicated number of people in AO, AC, and CO projects (the overlap count calculated in step 1 prior to dividing by 3) and the duplicated count of people in AO and AC projects (the overlap count calculated in step 2 prior to dividing by 2) from the duplicated count of people served in multiple household types (also calculated in step 2). This number is then multiplied by the percentage of duplicated people in either (a) AO and CO projects or (b) AC and CO projects who are in AO and CO projects. As with CoCs with no AC projects, we divide this number by 2 to arrive at the estimated unduplicated number of people in both AO and CO projects.

4. **Calculate the number of people that only overlap both participating AC and CO household-type projects.** If the CoC has AC and CO projects but no AO projects, the overlap is calculated by multiplying (a) the number of duplicated people in AC or CO projects by (b) the percentage of duplicated people in AC or CO projects that are in both AC and CO projects. We divide this number by 2 to arrive at the estimated unduplicated number of people that are in both AC and CO projects. If the CoC has AO, AC, and CO projects, the number of people overlapping AC and CO projects is determined by first calculating the duplicated count of people in AC and CO projects. To do this, we subtract the duplicated number of people in AO, AC, and CO projects (the overlap count calculated in step 1 prior to dividing by 3), the duplicated count of people in AO and AC projects (the overlap count calculated in step 2 prior to dividing by 2), and the duplicated count of people in AO and CO projects (the overlap count calculated in step 3 prior to dividing by 2) from the duplicated count of people served in multiple household types (calculated in step 2). As with CoCs with no AO projects, we divide this number by 2 to arrive at the estimated unduplicated number of people in both AC and CO projects.

5. **Calculate the number of people in participating AO Only, AC Only, and CO Only household-type projects by excluding those in multiple/overlapping household types.**

    a. To calculate the number of people in participating projects that are only in the AO household type only, we first calculate the total number of people in participating projects who are in the AO household type and any other household type. To do this, we sum the number of people in (a) participating AO, AC, and CO household-type projects (calculated in step 1), in (b) participating AO and AC household types only (calculated in step 2), and (c) participating AO and CO household types only (calculated in step 3). We then subtract this number from the number of people in the AO household type to arrive at the number of people in participating projects that are only in AO.

    b. To calculate the number of people in participating projects that are in the AC household type only, we first calculate the total number of people in participating projects who are in the AC household type and any other household type. To do this, we sum the number of people in (a) participating AO, AC, and CO household-type projects (calculated in step 1), in (b) participating AO and AC household types only (calculated in step 2), and (c) participating AC and CO household types only (calculated in step 4). We then subtract this number from the number of people in the AC household type to arrive at the number of people in participating projects that are in AC only.

    c. To calculate the number of people in participating projects that are only in the CO household type, we first calculate the total number of people in participating projects who

are in the CO household type *and* any other household type. To do this, we sum the number of people in (a) participating AO, AC, and CO household-type projects (calculated in step 1), in (b) participating AO and CO household types only (calculated in step 3), and (c) participating AC and CO household types only (calculated in step 4). We then subtract this number from the number of people in the CO household type to arrive at the number of people in participating projects that are only in CO.

Part B. Estimate unduplicated count of household-type overlap for people in all projects, participating and non-participating, by CoC.

We use the information on the percent of people in each of the seven household-type combinations in participating projects within each CoC and our earlier imputed estimate of the duplicated number of people in AO, AC, and CO to estimate the number of unduplicated people in the seven household-type combinations in all projects (i.e., participating and non-participating) within the CoC.[20]

1. **Estimate household-type overlap among CoCs in which AO, AC, and CO projects are all 100 percent participating.** If AO, AC, and CO are all 100 percent participating (including CoCs that serve only two household types and have 100 percent reporting in those two household types), the estimated total number of people in each household type (AO Only, AC Only, CO Only, AO and AC Only, AO and CO Only, AC and CO Only, and AO, AC, and CO) is equal to the participating number of people in each household type calculated above. That is, we use the actual observed number.[21]

2. **Estimate overlap among CoCs in which at least 2 of the 3 household types have at least 50 percent HMIS participation.[22]**

    a. Estimate the total number of people in participating and non-participating projects that overlap in all three (AO, AC, and CO) household types. To do this, we first estimate the total number of unduplicated people (in participating and non-participating projects) that overlap in the two household types with at least 50 percent participation (including those that overlap in only those two household types and those that overlap in all three household types). To estimate this unduplicated count, we multiply (a) the total duplicated number of people in either of the two 50 percent plus household types in participating and non-participating projects by (b) the percentage of duplicated people in participating projects in either of these two household types that are in both household types. We divide this number by 2 to arrive at the estimated unduplicated number of people that are in both household types (including those that are also in the third household type) in participating and non-participating projects.[23] Then estimate the total number of overlapping people in all three household types—AO, AC, and CO—by multiplying (a) the number of unduplicated people (in participating and non-participating projects) in both 50 percent plus household types (including those in all three household

[20] If the estimated total number of people across all projects is smaller than the calculated number of people in participating projects from Part A, the participating count is used instead.
[21] By extension, if a CoC only serves one household type, overlap would be 0.
[22] There is one exception. If participation is above 50 percent but less than 100 percent in all three household types, household-type overlap is calculated using the same method that was applied for the participating portion of CoCs (Part A above).
[23] If participation is 100 percent in the two household types with greater than 50 percent participation, the unduplicated count is instead equal to the number of participating people in both 50 percent plus household types (including those in all three household types).

types) by (b) the percentage of participating unduplicated people in both 50 percent plus household types who are in all three household types.[24,25]

b. Estimate the total number of people in participating and non-participating projects that only overlap the two household types with at least 50 percent participation (including CoCs that serve only two household types and have at least 50 percent reporting in those two household types). This estimate is derived by subtracting the number of overlapping people in all three household types calculated in step 2a from the number of unduplicated people in both 50 percent plus household types.

c. Estimate the total number of people in participating and non-participating projects that only overlap the first household type with at least 50 percent participation and the household type that may not have 50 percent or more participation. To do this, we first estimate the total number of duplicated people (in participating and non-participating projects) in the household type with more or less than 50 percent participation and either the first or second household types with 50 percent plus participating projects. To estimate this duplicated number, we multiply (a) the total duplicated number of people in any of the three household type categories by (b) the percentage of duplicated people in participating projects in any household type combination who are in the household type with more or less than 50 percent participation and either the first or second household type with 50 percent plus participating projects. [26] Then estimate the total number of duplicated people in the first household type with at least 50 percent participation and the household type that may not have 50 percent or more participation by multiplying (a) the number of duplicated people (in participating and non-participating projects) in the household type with more or less than 50 percent participation and either the first or second household types with 50 percent plus projects by (b) the percentage of duplicated people in participating projects with more or less than 50 percent participation and either the first or second household types with 50 percent plus projects who are in the household type with more or less than 50 percent participation and the first household type with 50 percent plus projects. We divide this number by 2 to arrive at the estimated unduplicated number of people that are in the first household type with at least 50 percent participation and the household type that may not have 50 percent or more participation.

d. Estimate the total number of people in participating and non-participating projects that only overlap the second household type with at least 50 percent participation and the household type that may not have 50 percent or more participation. To do this, we first estimate the duplicated number of total people (in participating and non-participating projects) in the second household type with at least 50 percent participation and the

[24] If participation is 100 percent in two of three household types and at least 50 percent in the third household type, the percentage is at the CoC level. If participation is less than 50 percent in one of the household types, the national percentage is used.

[25] For all estimates of overlap between household types, the overlap cannot be larger than the number of people in any of the overlapping household types. If the overlap is larger, the minimum of these household types is used instead. For example, if the calculated overlap between AO, AC, or CO household types is larger than the number of people in AO, AC, or CO households, the overlap is set to the minimum value of AO, AC, or CO households.

[26] Here and for all similar steps when applying percentages to duplicated counts, if participation is 100 percent in two of three household types and at least 50 percent in the third household type, the percentage is at the CoC level. If participation is less than 50 percent in one of the household types, the national percentage is used.

household type that may not have 50 percent or more participation. To estimate this duplicated count, we subtract the duplicated number of people in both the first household type with 50 percent plus participation and the household type with more or less than 50 percent participation (the final estimate from Step 2c but multiplied by 2 to make duplicated) from the duplicated number of people in the household type with more or less than 50 percent participation and either the first or second household types with 50 percent plus participating projects (also estimated in Step 2c). We divide this number by 2 to arrive at the estimated unduplicated number of overlapping people in the second household type with at least 50 percent participation and the household type that may not have 50 percent or more participation.

e. Estimate total people in AO Only, AC Only, and CO only projects excluding those in multiple/overlapping household types. This estimation uses the same method as calculating the number of people in participating projects in these household types as described in Part I section 5, but for total projects instead of participating projects.

3. **Estimate overlap among CoCs in which fewer than two household types have at least 50 percent HMIS participation.[27]**

   a. As in Step 2, we use the data on participating projects to calculate the percentage of duplicated people that are in each of the seven household type combinations. We then apply these percentages to the imputed estimate of the duplicated number of people in AO, AC, and CO in all projects to estimate the number of unduplicated people in the seven household-type combinations in all projects (participating and non-participating) within the CoC. Due to the lower participation rate for these CoCs, these percentages are applied at the national level rather than at the individual CoC level.

   b. The step-by-step procedures used to estimate the number of people in all projects in each household-type combination is the same as in Part A, the method used for calculating overlap among people in participating projects.[28] The only departure from Part A is that total counts instead of participating counts are used in the estimations.

Finalizing Weights

Constructing Weighting Classes

The goal in constructing weighting classes is to choose classes in which units within each class are more similar to each other than to units in other classes. Within each of the 21 project-household type combinations,[29] the weight for each weighting class is defined as the estimated unduplicated number of

[27]This procedure includes CoCs that serve only two household types and have fewer than 50 percent reporting in at least one of those two household types.

[28]If the estimated total count is smaller than the calculated participating count, the participating count is used instead.

[29] There are three project types: (1) Emergency Shelter/Safe Haven/Transitional Housing, (2) Permanent Supportive Housing, and (3) Rapid Re-Housing. In each project type, there are seven household type combinations: (1) Adults Only, (2) Adults and Children Only, (3) Children Only, (4) combination of Adults Only and Adults and Children, (5) combination of Adults Only and Children Only, (6) combination of Adults and Children and Children Only, and (7) combination of Adults Only, Adults and Children, and Children Only.

people in participating and non-participating projects in the weighting class divided by the unduplicated number of people in participating projects in the weighting class.

Weight classes were constructed using propensity stratification separately for each of the 21 project-household groups. With propensity stratification, CoCs were grouped into a weighting category with other CoCs with propensity scores most similar to their own. These propensity scores were generated by modeling a response outcome on a variety of characteristics, including percentage of persons in each geographic category,[30] percentage of persons in each race category,[31] percentage of persons who were veterans, youth, or chronically homeless, percentage of persons who were HIV positive, and percentage of persons who received McKinney-Vento funding.[32] Two versions of outcomes for propensity stratification were tested, a continuous measure of HMIS participation and a binary measure of HMIS participation. The continuous measure models the proportion of the CoC that is participating while the binary measure treats the CoC as participating when any portion of the CoC is participating. For each outcome measure, a variety of model types were tested. One model included all possible predictor variables (covariates) while other models limited predictors to those either chosen from automated selection models or tested manually.

For each of the 21 project-household groups, the stratification scheme[33] and model type (covariate combination) selected was the least likely to have less than two participating CoCs or an outlier weight (set as greater than 20), had the smallest difference between the weighted participating count and unweighted total count across all characteristics,[34] and prioritized binary measures of participation and weighted estimates.

Computing Final Weights

Two sets of weights are applied, national person weights and national household weights.

As described above in the construction of weight classes, the national person weights are computed as the inverse of the participation rate (the estimated total count of persons divided by the reported count of persons) in each weighting class.

The national household weights are stratified by the 3 broader household types (Adults Only, Adults and Children, and Children Only). These household weights are defined as the ratio of the weighted household counts (calculated from the national person weights applied to unweighted person counts) to the unweighted household counts.

---

[30] Geographic categories are defined as urban, suburban, rural, or unknown.

[31] Percentages of CoCs in each race category by household type originated from Point in Time homeless data. These percentages were drawn from sheltered homeless data in 2022 if the total sheltered count in that year was greater than zero for that household type. If the sheltered count was not greater than zero but the total homeless count was greater than zero for that household type in 2022, these race percentages were based on all homeless persons in 2022. If the total homeless persons for that household type in 2022 was not greater than zero but the sheltered count in 2021 for that household type was greater than 0, these percentages were based on the sheltered homeless in 2021. If the sheltered count in 2021 for that household type was not greater than 0 but the sheltered count in 2022 across all household types was greater than 0, these percentages were based on the sheltered count in 2022 across all household types. If the sheltered count in 2022 across all household types was not greater than 0 but the total homeless count across all household types in 2022 was greater than zero, these percentages were based on the total homeless count in 2022 across all household types. If the total homeless count across all household types in 2022 was not greater than 0, we followed the same sequence using earlier years. The same race percentages were used across project type and participation status.

[32] Characteristics were excluded from the model if they were not applicable for the particular project type and household combination. For example, percentage of CoCs who are veterans are not applicable to child-only household types.

[33] Stratification schemes tested were propensity scores grouped by deciles, by quintiles, and a variety of other groupings.

[34] This is the complete set of characteristics included as predictor variables from the model that estimated propensity scores

Applying Final Weights

The calculated weights are applied to each CoC's reported counts. To estimate national person counts, reported person counts (on person characteristics such as age, gender, or race) were multiplied by national person weights. To estimate national household counts, reported household counts (on household characteristics such as household size and household geography) were multiplied by national household weights. Weights were applied separately by weighting class within project-household groups.

To calculate the 95% confidence intervals across all reporting categories and within household types, the standard error was multiplied by 1.96 and then subtracted from and added to the weighted estimates. The standard error calculation was derived from the same strata used for constructing the sampling weights and the sampling rate.

Table 7. 95% Confidence Intervals of Sheltered Homelessness by Household Type and Program Type, 2022

	Weighted Estimate	Standard Error	95% Confidence Interval (Lower)	95% Confidence Interval (Upper)
Emergency Shelter, Transitional Housing, or Safe Havens: Total People (active in residence)	1,388,425	62,681	1,265,413	1,511,438
Emergency Shelter, Transitional Housing, or Safe Havens (active in residence): Total Households	1,066,514	66,418	936,099	1,196,929
Permanent Supportive Housing: Total People (active in residence)	387,694	23,422	341,682	433,707
Permanent Supportive Housing (active in residence): Total Households	292,501	21,620	250,019	334,983
Rapid Rehousing: Total People (active in residence)	256,653	13,472	230,200	283,105
Rapid Rehousing (active in residence): Total Households	147,468	8,246	131,272	163,665

Additional Data and Method Notes

The study team used data from the ACS Census Public Use Microdata Sample (PUMS) file produced by the Census Bureau. The report includes data from the ACS to show estimates of the total U.S. population and the population living in poverty in the U.S. to contextualize the population experiencing homelessness in the U.S. The 2022 ACS 1-year estimates were used for comparison with 2022 HMIS-based data.

The ACS PUMS identifies the Region, Division, State and Public Use Microdata Areas (PUMAs) where a person or housing unit record is located. PUMAs are the most detailed geographic area available in the ACS PUMS. PUMAs are non-overlapping areas that partition each state into areas containing approximately 100,000 residents. In addition to ACS PUMS, the study team used the PUMA land use crosswalk to determine the geography type (urban, suburban, or rural) at the project level. To determine geography type for each PUMA, PUMA boundaries were laid over land use GIS layers, processed from the National Center for Education Statistics Locale Boundaries data. Calculations of population by land use within each PUMA were conducted and assigned geography type with the highest population percentage. The population land use estimates are based on Census 2020 decennial data at the census block level. While the PUMAs within the boundaries and density requirements of an urban, suburban, or rural area were easily coded, those that overlap with other geography types required adjustments.