Journal of Urban Economics 136 (2023) 103559

Contents lists available at ScienceDirect

Journal of Urban Economics

journal homepage: www.elsevier.com/locate/jue

# The size and Census coverage of the U.S. homeless population ☆

Bruce D. Meyer [a], Angela Wyse [b,*], Kevin Corinth [b,c]

[a] University of Chicago, NBER and AEI, United States
[b] University of Chicago, United States
[c] American Enterprise Institute, United States

ARTICLE INFO

*Keywords:*
Homelessness
American community survey
Decennial census
HUD PIT count

ABSTRACT

Fundamental questions about the size and characteristics of the homeless population are unresolved because it is unclear whether existing data are sufficiently complete and reliable. We examine these questions and the coverage of new microdata sources that are designed to be nationally representative. We compare two restricted data sources largely unused to study homelessness, the 2010 Census and American Community Survey (ACS), to restricted Homeless Management Information System (HMIS) data, HUD's public-use point-in-time (PIT) estimates, and the Housing Inventory Count (HIC) at the national and individual level. We also develop a new approach to estimating the size of the sheltered homeless population using linked Census and HMIS microdata. Our analyses suggest that on a given night there are about 400,000 people experiencing homelessness in shelters in the U.S. and about 200,000 people sleeping on the streets, with this latter estimate subject to greater uncertainty. More than 90 percent of those in shelters appear to be counted in the Census, although many are classified as housed or in other group quarters, due largely to ambiguity in the definition of a homeless shelter. This paper lays the foundation for pathbreaking future work with these data on the U.S. homeless population.

## 1. Introduction

Despite widespread concern about those experiencing homelessness, many of the most basic questions about this population, including the first-order question of population size, are unresolved. Relatedly, the extent to which the Decennial Census and Census Bureau surveys include those experiencing homelessness is unclear in Census documentation and publications, and the empirical extent of coverage has not been examined. In this paper, we compare two restricted data sources that have been largely unused to study homelessness to administrative shelter records and less detailed public data. We also develop a new approach to estimating the size of the sheltered homeless population by linking together Census and administrative shelter microdata, an approach that under our stated assumptions provides a reliable estimate of the true population. We evaluate the usefulness of these datasets to advance our understanding of this difficult-to-study group and lay the foundation for pathbreaking future work using these data.

Efforts to count the U.S. homeless population confront substantial challenges. Because people experiencing homelessness lack a fixed domicile, they cannot be counted using standard address list-based approaches like those most often used in the Census and household surveys. They must instead be counted in the shelters, soup kitchens, encampments, vehicles, or parks where they happen to be staying at a given time. This difficulty is at times compounded by mistrust of authorities, mental illness or substance abuse, involvement in the underground economy, local ordinances that restrict activities associated with homelessness, or other factors that contribute to a desire not to be found (Corinth, 2015; Glasser et al., 2013).

Given these difficulties, the reliability of available estimates, particularly the Department of Housing and Urban Development (HUD)'s point-in-time (PIT) count, is frequently called into question. The PIT is widely cited in the media and often used to allocate resources and inform policy, yet the handful of existing studies on its quality have been limited in geography and scope and are outdated

☆ The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applies to this release, authorization numbers: CBDRB-FY20-ERD002-004, CBDRB-FY2022-CES005-006, and CBDRB-FY2022-CES005-008. We thank the U.S. Census Bureau for their support and thank John Abowd, Mark Asiala, George Carter, James Christy, Dennis Culhane, Kevin Deardorff, Conor Dougherty, Ingrid Gould Ellen, Anne Fletcher, Katie Genadek, Kristin Kerns, William Koerber, Margot Kushel, Larry Locklear, Tim Marshall, Brian McKenzie, Brendan O'Flaherty, James Pugh, Trudi Renwick, Annette Riorday, Nan Roman, William Snow, Eddie Thomas, Matthew Turner, and John Voorheis for providing feedback and answering our questions and Gillian Meyer, Connor Murphy, and Sophie Yang for research assistance. We appreciate the financial support of the Alfred P. Sloan Foundation, the Russell Sage Foundation, the Charles Koch Foundation, the Menard Family Foundation, and the American Enterprise Institute.

* Corresponding author.
*E-mail address:* awyse@uchicago.edu (A. Wyse).

https://doi.org/10.1016/j.jue.2023.103559
Received 26 July 2022; Received in revised form 2 May 2023

(Hopper et al., 2008; Agans et al., 2014). A 2020 report from the U.S. Government Accountability Office (GAO) determined that the PIT "did not provide a reliably precise estimate of the homeless population," in part, according to the report, because of the decentralized and non-uniform way that local bodies carry out their counting operations. O'Flaherty (2019) observes that PIT data on the unsheltered homeless population are largely gathered by a "loosely supervised army of amateur volunteers" whose "diligence, understanding of the process, and lack of bias are all open to question." The completeness and coverage of shelter-use microdata, which are employed in the PIT's sheltered homeless estimates, have gone largely unstudied. By comparing the PIT's estimate of the U.S. homeless population to independent estimates, this paper provides the most comprehensive assessment to date of the quality of both the aggregate PIT and the microdata underlying its sheltered population estimates.

Our approach draws on restricted microdata from the 2010 Census, the American Community Survey (ACS), and Homeless Management Information System (HMIS) databases from Los Angeles and Houston. The ACS and HMIS include people in homeless shelters, while the Census includes both sheltered and unsheltered homeless individuals. We compare these restricted data to each other and to HUD's PIT estimates and the Housing Inventory Count (HIC). Our restricted data have important advantages over public data. Like the PIT, the ACS and Census are designed to be representative of the entire U.S. homeless population. Unlike the PIT, however, the Census, ACS, and HMIS include individual linkage keys so that the microdata can be linked across sources and to administrative data to examine longitudinally a range of social and economic characteristics. The ACS and HMIS data also in themselves contain a rich set of information about homeless individuals. By examining the coverage and reliability of Census, ACS, and HMIS data, this paper lays the foundation for future work taking full advantage of these datasets to learn about the U.S. homeless population. This paper also provides valuable insight into the coverage of people experiencing homelessness in the Census and household surveys, some of the most fundamental sources of data on the U.S. population.

We begin with an aggregate comparison of unsheltered and sheltered homeless estimates in the Census and PIT. We find that the Census and PIT's unsheltered estimates are quite close to one another, providing encouraging but not definitive evidence of the estimates' accuracy. Moreover, despite what appear at first to be major differences in sheltered homeless estimates, the Census and PIT in fact produce similar estimates once we account for straightforward definitional and weighting differences. Specifically, the PIT's sheltered homeless population estimate includes people in domestic violence shelters, those in voucher-funded hotel and motel rooms, and people in non-shelter facilities, whereas the Census and ACS classify these groups of people as belonging to other, non-homeless statuses. We also describe an aspect of the ACS's weighting methodology that inflates sheltered homeless population estimates by over 30 percent in each year to represent people not included in the survey's scope. Adjusting for straightforward definitional differences and correcting the ACS weighting brings the Census and ACS estimates much closer to the sheltered PIT estimate. The fact that these two sources produce similar estimates despite employing substantially different methods bolsters our confidence in both estimates, although we discuss potential sources of bias relative to the true homeless population that may net out in aggregate comparisons.

Our second set of analyses compare data sources at the person level. We link HMIS shelter use microdata from Los Angeles and Houston to the 2010 Census to learn more about both sources' coverage and to assess the usefulness of Census microdata to study this population. Under stated assumptions and after accounting for likely errors in shelter exit date reporting in the HMIS data, we estimate that about 80–95 percent of people who were indicated as being in HMIS shelters on the date of the Census's homeless counting operation were counted in the Census, although only about 35–45 percent of them were included in the Census's sheltered homeless count, with the rest being counted as housed, unsheltered homeless, or in other types of group quarters facilities. We provide evidence that errors in shelter exit date tracking in HMIS are an important reason for these status discrepancies. We also show that many HMIS facilities, particularly transitional shelters where homeless individuals can reside for up to two years, appear to have been often classified as housing units or other types of group quarters rather than homeless shelters by the Census. Finally, we note that many people may have responded to the Census while housed before entering a shelter or after exiting it during the long window of potential Census response, which ran from mid-March to well into May 2010.

Unexpectedly, our microdata comparisons reveal extensive double-counting of homeless individuals in the 2010 Census. We estimate that 21–24 percent of the sheltered homeless, 45–56 percent of those counted in soup kitchens and while using food vans, and 29–35 percent of those at outdoor locations had at least one housed record in addition to their homeless record in the 2010 Census. We rule out widespread erroneous linkages and misclassification of housed people as homeless and provide evidence that double counting arises primarily when homeless individuals are included on the Census questionnaire of a household where they occasionally reside or where they resided within a few months of the Census's homeless counting operation.

Finally, we develop a new approach to estimating the size of the sheltered homeless population using linked Census and HMIS shelter microdata. This method draws on dual system estimation techniques used frequently in demography and in ecology and allows us to obtain a reliable estimate of the true population under certain assumptions. In brief, we take the share of people in HMIS shelters in Los Angeles and Houston on the Census date who were included in the Census's homeless counting operation as an estimate of the share of the true sheltered homeless population in the Census. We then scale up the Census estimate by the inverse of this share to adjust for under coverage and obtain an estimate of the true sheltered homeless population. This approach does not make assumptions about the completeness of the Census or PIT, but does rely on several assumptions, including the assumption that those counted and uncounted in the Census are equally likely to appear in the HMIS data, an assumption that is plausible but difficult to verify. Using these methods, we estimate the sheltered homeless population size in 2010 to be 367,000–382,000 people, or about 5–10 percent lower than the 2010 PIT estimate and about 27–32 percent larger than the Census count after straightforward definitional adjustments. These analyses suggest that about 93–97 percent of people who were in shelters on the Census date were included in the Census in some status. In addition to providing a new population estimate, this section serves as a blueprint for future researchers seeking to estimate the homeless population as additional data become available.

Our analyses produce several key insights into the size of the U.S. homeless population. We find that, despite what initially appear to be substantial differences between 2010 Census, ACS, and PIT estimates of the homeless population, these sources produce very similar estimates once we account for definitional and weighting differences. We evaluate these aggregate comparisons for the sheltered homeless population with our dual system approach. Taken together, these estimates suggest that on a given night there are about 400,000 people experiencing homelessness in shelters in the U.S. and about 200,000 people sleeping on the streets, with the latter number subject to greater uncertainty. At the same time, our results highlight the fact that there is considerable ambiguity about what types of facilities constitute homeless shelters and that population estimates are sensitive to how these ambiguities are resolved.

Beyond population estimates, this paper also advances our understanding of homeless individuals' coverage in the Census. Our findings suggest that the Census was able to include more than 90 percent of sheltered homeless individuals, although oftentimes it classified them as housed or as residing in non-shelter group quarters facilities. At the same time, widespread instances of double counting of homeless individuals in the Census paint a picture of a highly mobile population that frequently transitions between housed and homeless living situations. These findings suggest that household surveys that rely on Census address lists may incorporate homeless individuals more often than previously thought. By establishing the broad coverage and reliability of the new data sources, this paper lays the foundation for pathbreaking future work using the Census, ACS, and HMIS datasets, including efforts to learn about this population's longitudinal patterns of income and safety net participation and the heightened mortality risk associated with homelessness.

This paper proceeds as follows. Section 2 discusses past efforts to estimate the size of the homeless population and summarizes the literature on the quality of available estimates. We also define homelessness and discuss the merits of the definition we use relative to others. Section 3 describes our data, including the 2010 Census, ACS, PIT, and related datasets. Sections 4 and 5 describe our methodology and results for aggregate and microdata comparisons, respectively. Section 7 describes our dual system estimate of the sheltered homeless population size. Section 8 discusses these findings and Section 9 concludes.

## 2. Background and related literature

### 2.1. Prior efforts to estimate the homeless population size

In the 1980s, an apparent rise in homelessness and a surge in media coverage inspired numerous attempts to estimate the U.S. homeless population. Intense controversy surrounded these efforts from the beginning. HUD's first national estimate in 1984 placed the population between 250,000 and 350,000, but their findings were criticized by advocacy groups who maintained that the true number was as high as three million (U.S. General Accounting Office 1985). In a 1992 meta-analysis, Shlay and Rossi (1992) observed that most of the 60 studies they reviewed relied on an unreasonable degree of extrapolation or speculative assumptions and amounted to "sheer guesses" of the homeless population size.

HUD began publishing point-in-time (PIT) estimates in its Annual Homeless Assessment Report (AHAR) in 2007 in response to a directive from Congress. As a national source of longitudinal population estimates, the PIT represents a major advance over previous efforts to count the homeless. It is nevertheless imperfect. HUD engages local homeless service coordinating bodies, known as Continuums of Care (CoCs), to carry out PIT operations and allows them to employ a range of methods. In practice, the techniques used and resources invested vary substantially – as does, presumably, the quality of estimates (U.S. Department of Housing and Urban Development 2014).[1]

A small body of research examines the completeness of unsheltered PIT counts. Several studies have dispatched decoy homeless individuals on the night of the PIT and later reported the share that were included in the PIT. One such study during a 2005 point-in-time count in New York City found that 30 percent of decoys were missed by enumerators (Hopper et al., 2008). The authors also surveyed a sample of homeless individuals about their sleeping arrangements the night of the PIT and estimated that 31–41 percent would not have been visible to counters. In Los Angeles in 2009, Agans et al. (2014) conducted a post-PIT telephone survey asking residents if they knew of homeless individuals who had spent the previous night on private property and would have therefore been missed by that city's PIT. The authors estimated that 20 percent of Los Angeles's unsheltered homeless population would have been missed by the PIT.

The literature pays less attention to the sheltered PIT. These estimates are thought to be more reliable because they are in many cases derived from the Homeless Management Information System (HMIS) database. In practice, HMIS data quality varies between shelters and over time. Cronley (2011) found wide variation in the frequency and thoroughness of HMIS record-keeping among 24 homeless service provides in Michigan and Tennessee during the early years after the system's implementation.

The Census made its first systematic attempt to enumerate homeless individuals during a 1990 operation called Shelter and Street Night (S-Night). S-Night's count of 228,621 individuals fell far below consensus estimates at the time, prompting the Census Bureau to state that "S-Night was not intended to, and did not, produce a count of the 'homeless' population of the country" (Martin 1992). Various S-Night evaluations found that decoys deployed in five cities to act as unsheltered homeless persons were only counted 22 to 66 percent of the time (Wright and Devine 1992).

The Census Bureau aimed to improve on the S-Night methodology with its first Service-Based Enumeration (SBE) in 2000, visiting shelters, food vans, soup kitchens, and a list of pre-identified outdoor locations. This effort produced a count of 280,527 individuals and again received an official caveat: "We cannot be certain that all places were covered or that all people normally using shelters were included in the shelter counts. Nor can our coverage of targeted outdoor locations be considered to have been exhaustive due to the difficulties in mapping such temporary and elusive sites" (Smith and Smith 2001).

The 2010 SBE fared better than the previous two attempts. Meyer et al. (2022) provide a preliminary analysis of the characteristics of those included in the 2010 Census homeless counting operation and demonstrate the types of analyses that can be undertaken once the coverage of this population in the Census is better understood. We discuss the 2010 SBE in depth in Section 3.2 of this paper.

### 2.2. Defining the homeless population

In this paper, we follow HUD's definition of literal homelessness. People are literally homeless if they have "a primary nighttime residence that is a public or private place not designed for or ordinarily used as a regular sleeping accommodation for human beings, including a car, park, abandoned building, bus or train station, airport, or camping ground" (the unsheltered) or if they are living in "a supervised publicly or privately operated shelter designated to provide temporary living arrangements (including congregate shelters, transitional housing, and hotels and motels paid for by charitable organizations or by federal, State, or local government programs for low-income individuals)." This is the definition of homelessness that guides HUD's point-in-time count and it aligns closely with the population targeted by the Census's homeless counting operation.

We distinguish people experiencing literal homelessness from those who are precariously housed, have low-quality accommodations, or face imminent risk of homelessness for some other reason. Policymakers and researchers are often rightly concerned about hardships faced by people in these categories and at times include them in official definitions of homelessness. The Department of Education, for example, defines homelessness to include children "sharing housing with others due to loss of housing, economic hardship, or a similar reason," otherwise known as doubling up (U.S. Department of Education, 2021).

While such situations often reflect housing-related hardship, we maintain that literal homelessness is the most useful definition for economists. For one thing, literal homelessness indicates a level of material deprivation that in most cases exceeds the hardship experienced by those who are precariously housed or doubled up. The choice of where to live reflects a complex economic calculation by maximizing agents whose choice set typically includes homeless shelters. When shelter beds are available, the decision to share housing or live in subpar accommodations indicates a revealed preference for these living arrangements over literal homelessness.

Moreover, it is not clear that shared housing reflects economic hardship in most cases. There are many reasons why shared housing might be

[1] The 2009 AHAR, for example, singled out Detroit and New Orleans as having conducted counts of particularly suspect quality that year (U.S. Department of Housing and Urban Development 2010).

preferable to solo living options, as is well documented in the household formation literature. Reasons include the sharing of quasi-public goods like appliances, bathrooms, and living space and facilitating trades of time, resources, and services like housework or informal caregiving for children or the elderly (Browning et al., 2014). Because it is voluntary, the decision to share living quarters should not be a priori thought of as bad.

As a practical matter, existing data do not allow researchers to identify people for whom shared accommodations reflect extreme hardship. Such a determination would require detailed knowledge of all options in the agent's choice set, including the quality of accommodations, precariousness of tenure, and other factors that could make housing alternatives extremely undesirable (e.g. abuse or neglect at home or unsafe conditions in shelters). For example, when the Department of Education trains educators to identify children who qualify for homeless services due to doubling up, it instructs them to interview parents and/or students extensively to determine whether personal housing is available, whether they left their last housing situation under duress (e.g. were evicted or fled abuse or neglect), and whether their shared housing meets the subjective criteria of being "fixed, regular, and adequate" (U.S. Department of Education, 2021). Educators then make determinations of doubled-up homelessness on a case-by-case basis. As these training materials illustrate, the information requirements for making such a determination go far beyond the questions asked in household surveys.

### 2.3. Time-frame considerations in defining homelessness

We emphasize estimates of the number of people who are homeless at a point in time in this paper. This decision reflects, in part, the availability of comparable estimates in different data sources. While HUD produces estimates of the number of people who used homeless shelters each year, these estimates are not available for the unsheltered and there are no comparable estimates for the sheltered in other data sources. Moreover, HUD's annual estimates are based on data collected by a subset of shelters and then extrapolated to the entire U.S. using assumptions that are difficult to validate.

Relative to interval-based population estimates, cross-sectional estimates include a greater share of people experiencing long-term or repeated homeless spells. This group likely includes people with exceptional difficulty maintaining housing as well as those who have secured extended shelter placements but nevertheless meet the definition of literal homelessness because of how HUD classifies those facilities. As discussed in O'Flaherty (2019), which temporal convention is most appropriate depends on the question at hand and our (as-yet very limited) understanding of how the social and private costs of homelessness vary with time spent homeless. We note, however, that the decision to emphasize the cross-sectional homeless population aligns with the approach used in other literatures, including those that study number of people who are in poverty or unemployed at a point in time.

## 3. Data

This section describes the five sources of data on the homeless population used in this paper: the 2007–2021 HUD PIT and the associated Housing Inventory Count (HIC) dataset, the 2010 Census, the 2006–2018 ACS, and the HMIS microdata from Los Angeles (2004–2014) and Houston (2004–2015).

### 3.1. HUD's point-in-time (PIT) estimates

HUD requires that CoCs produce sheltered homeless population estimates every year and unsheltered estimates at least every other year to maintain federal funding. CoCs' geographic areas can encompass a single city or county, a metro area, a collection of counties, or the so-called "balance of state" outside of one or two major cities. These estimates are known as the point-in-time (PIT) count because they count (or in most cases, estimate) the homeless population on a single night, typically in the last two weeks of January. Each CoC plans and executes its own counting operation using one or more of a set of HUD-approved methods, typically a combination of enumeration, surveys, and extrapolation, occasionally done with the help of outside consultants. Many CoCs rely on volunteers to conduct nighttime canvassing operations, while others conduct multi-day or morning after operations at service locations. CoCs attempt to mitigate double-counting of the same individual using various strategies – for example, by asking homeless individuals whether they have already been counted – but are limited in their ability to de-duplicate unsheltered individuals because they rarely collect identifying information. Sheltered counts often rely, at least in part, on extrapolation from Homeless Management Information System (HMIS) databases.

CoCs also compile an inventory of all beds available for occupancy on the night of the PIT each year. This inventory is published in a separate dataset called the Housing Inventory Count (HIC), which lists the number of beds available on the PIT date, the number of people sleeping there, the target population (e.g., veterans, domestic violence victims, people with HIV/AIDS), and the bed type (e.g., in a shelter, in a non-shelter location, or in the form of vouchers for hotels or motels).

### 3.2. 2010 Census Service-Based Enumeration (SBE)

The 2010 Census counted people experiencing homelessness during its Service-Based Enumeration (SBE) operation. Field staff visited emergency and transitional shelters, soup kitchens, food vans, and targeted non-sheltered outdoor locations (TNSOLs, e.g. street intersections or parks where homeless individuals were known to sleep) between March 29 and 31, 2010. The list of shelters and unsheltered locations was built using past Censuses' lists, internet research, and input from local and state governments homeless advocacy organizations, in addition to several advance visit and validation operations. Unlike the PIT, the Census trained enumerators to use uniform methods and apply the same standards nationwide when counting people experiencing homelessness. They also collected name and date of birth when possible.

The Census took several steps to ensure that the same individuals were not counted in multiple locations (Russell and Barrett 2013). People counted at soup kitchens and food vans were asked whether they had a usual home elsewhere and to provide an address. The Census later used a matching algorithm and clerical review to check whether the person was counted at that address and, if so, kept only the housed record. The Census also used this algorithm to de-duplicate person records within the SBE universe. However, the Census did not resolve potential duplicates between homeless shelters and housed or group quarters locations.

A team of non-Census researchers concluded that "there was a high level of cooperation between the homeless service providers such as shelter and day center administrators and the U.S. Census" (Glasser et al., 2013). Nevertheless, the Census Bureau has issued several caveats on the completeness of the SBE's homeless count. An official report noted that "people experiencing homelessness [could] be counted and included in the census via various operations [other than the SBE]," meaning that people in difficult-to-classify situations, such as those precariously housed with friends or acquaintances or residing in motels, might be grouped in with others who are not homeless in published counts (Smith et al., 2012).[2]

[2] For example, people residing temporarily in hotels, motels, campgrounds, or other transitory locations may have been counted during the Enumeration at Transitory Locations (ETL) operation, and the Census considers ETL facilities to be a housed status. Some definitions of homelessness also include people who are "doubled-up," i.e. sharing accommodations after losing prior housing or due to economic hardship. Such individuals would have been included on those households' housing unit questionnaires and not included in the SBE.

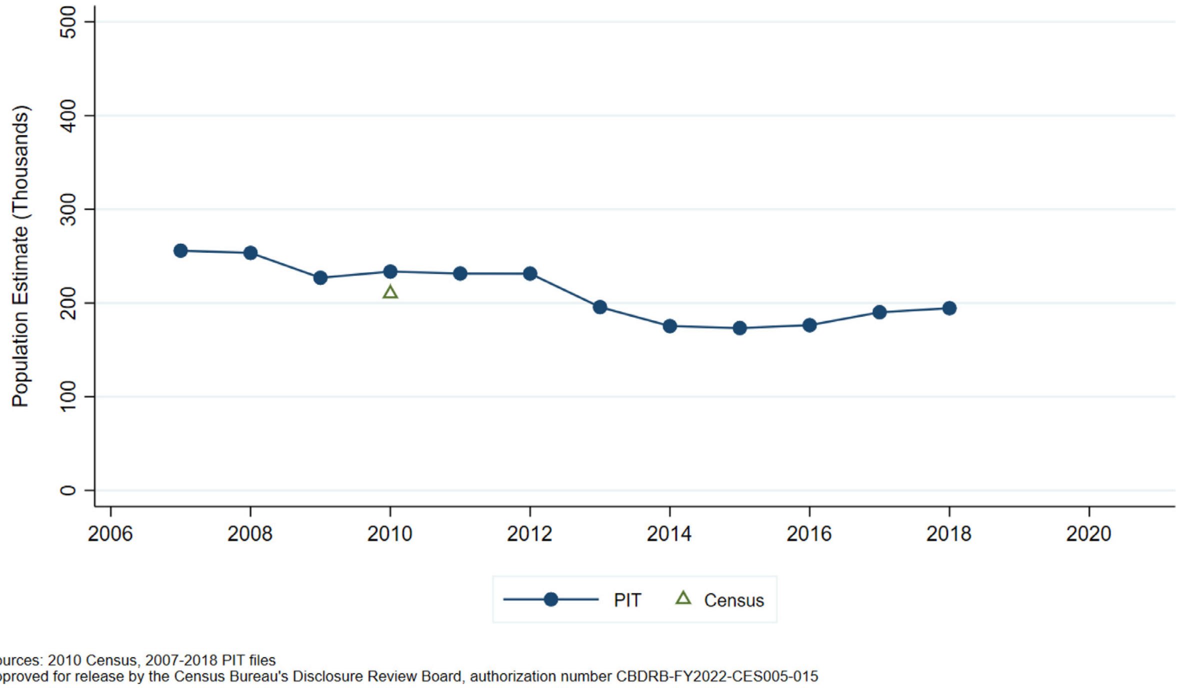


**Fig. 1.** Unsheltered homeless population estimates in the PIT and Census.

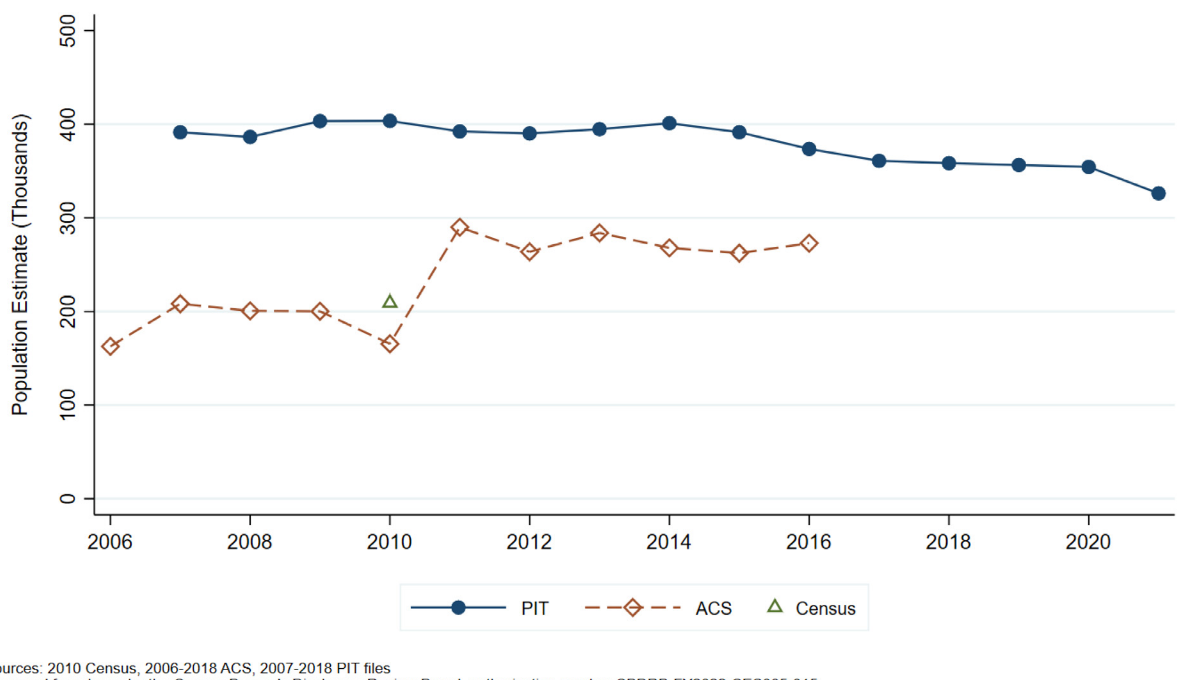


**Fig. 2.** Sheltered homeless population estimates in the PIT, ACS, and Census.

### *3.3. 2006–2016 American Community Survey (ACS)*

The ACS differs from the PIT and Census in that it only counts people experiencing homelessness in shelters, not those on the streets. It relies on random sampling and collects a much larger set of information than the other sources, including self-reported information on demographic characteristics, education, migration, and income and government program receipt. The ACS is conducted throughout the year and thus its population estimates approximate an annual average of point-in-time counts.

The ACS bases its sampling frame on extracts from the Master Address File, which is the Census Bureau's inventory of known housing units, group quarters (GQ) facilities like homeless shelters, transitory locations, and selected nonresidential units. Although the Census Bureau regularly updates this address file, the updating of GQ addresses between Censuses is operationally intensive and lags behind procedures for updating housing unit addresses (National Research Council 2012). As a result, the ACS's shelter inventory consists primarily of information from the most recent Census and likely becomes increasingly outdated in the ten years between Censuses. For example, of the homeless shelters selected for the 2008 ACS sample, about 42 percent no longer existed, were unoccupied, or had been converted into housing units (National Research Council 2012).

### 3.4. Homeless Management Information System (HMIS) data

In addition to the three sources of homeless estimates described above, this paper also draws on administrative shelter-use microdata from the Homeless Management Information System (HMIS) databases in Los Angeles (2004–2014) and Houston (2004–2015). Shelters that receive federal funding are required to track shelter use in an HMIS database, and some shelters that do not receive federal funding elect to do so as well.

Shelter administrators collect several data elements from all clients, including name and date of birth, social security number, and characteristics such as race, ethnicity, gender, veteran status, and disabling conditions. They also track the start and end dates of shelter enrollment and participation in some non-shelter programs like permanent supportive housing, rapid re-housing, and unsheltered outreach. Unlike Census data, HMIS data differentiate between emergency shelters and transitional housing and include shelter names. HMIS data are often used in part to generate CoCs' sheltered homeless PIT estimates, although HUD instructs CoCs to ensure that entry and exit date tracking is reasonably complete and accurate before relying on HMIS-based population totals in the place of canvassing or surveys administered on the night of the PIT operation (HUD 2012).

## 4. Comparisons of aggregate estimates

In this section, we compare aggregate sheltered homeless population estimates in the PIT with those in the Census and ACS. Our goal is to understand how much of the difference between sources can be attributed to straightforward definitional differences and weighting procedures. In doing so, we seek to make the Census and ACS estimates more comparable to the PIT as a precursor to other analyses. Although there are many ways to define homelessness, we make the PIT's definition our target because it is widely used by HUD and service providers.

Figs. 1 and 2 present estimates of the unsheltered and sheltered homeless populations for each year a given source is available.[3] In Fig. 1, we see that the 2010 unsheltered homeless population according to the PIT was 233,534, while the Census estimate was about ten percent lower at about 210,000. Fig. 2 shows that sheltered population estimates differ more substantially between sources. The sheltered population according to the PIT in 2010 was 403,543, while the Census estimate was 52 percent lower, at about 209,000. The ACS ranges from 41 to 54 percent of the PIT in the years 2006 through 2010 but then jumps to between 67 and 75 percent of the PIT in 2011 through 2016. This jump largely reflects the introduction of a new shelter list and the use of a new population benchmark after the 2010 Census rather than a change in the homeless population size.

### 4.1. Reconciling definitional differences between the PIT count and the Census and ACS

As a first step towards reconciling different estimates in the PIT count, Census, and ACS, we account for a handful of straightforward differences in the way these sources define homelessness. Specifically, the PIT's definition of sheltered homelessness includes people in several types of facilities outside the scope of the Census's Service-Based Enumeration and outside the scope of the ACS's sheltered homeless estimate, including domestic violence shelters, Safe Havens, voucher-funded hotel and motel rooms, and non-shelter facilities with beds for people experiencing homelessness. People residing in these facilities were included in the Census but classified as being housed or in other types of group quarters. For example, the Census classifies people in domestic violence shelters as being in religious group quarters and does not identify them separately even in restricted data to protect privacy. Safe Havens, which are small-scale facilities for individuals with a history of chronic homelessness and mental illness, are a form of supportive housing and hence classified as housing units in the Census.People residing in hotels and motels, while considered homeless by the PIT if their stays are funded by vouchers, would have been included in the Census during the enumeration of transitory locations, an operation that is separate from the SBE.[4] Beds in non-shelter facilities, which are included in the PIT, would not be included in the Census's SBE unless they had been identified during the Census's address list updating operation and validated as homeless shelters by a facility administrator.

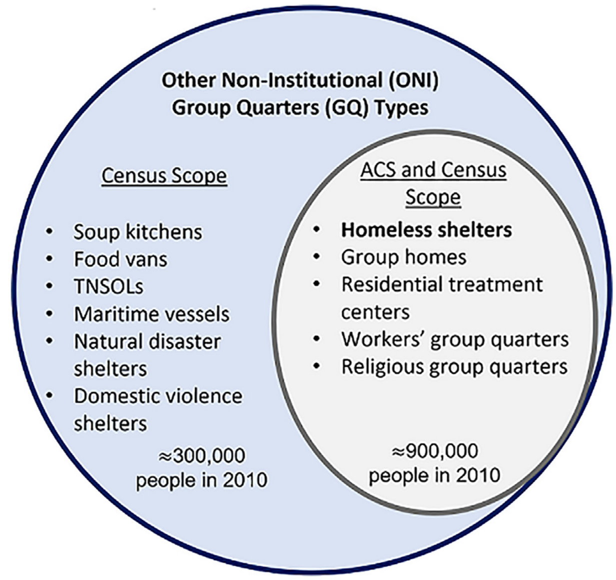


**Fig. 3.** Graphical representation of Other Non-Institutional (ONI) group quarters types.
Note: circles not to scale.

We adjust the aggregate Census and ACS estimates to better align their definition of sheltered homelessness with that of the PIT count. We obtain estimates of the number of people in Safe Havens from published HUD totals. For the other types of facilities, we can either directly calculate or estimate the PIT-only population using information available in the Housing Inventory Count (HIC)'s inventory of shelter beds. In some but not all years, the HIC includes each shelter's PIT count and indicators for whether the facility is a domestic violence shelter, whether it is voucher-based, and whether it is located in a non-shelter facility. For years where the HIC file is incomplete, or where a given data field is not available, we impute values using information in surrounding years.

### 4.2. Correcting bias from ACS weighting of the sheltered homeless

We next discuss an aspect of the ACS's weighting methodology that causes upward bias in its homeless population estimates. This bias arises from the ACS's use of population benchmarks in constructing person weights. Specifically, a final step of the ACS weighting methodology scales up person weights so that weighted population estimates match

[3] We exclude PIT and Census totals from U.S. territories in all of these analyses.

[4] Although the Census definition of emergency and transitional shelters technically includes "hotels and motels used to shelter people experiencing homelessness," in practice these sites would only be included in the SBE if a hotel or motel administrator told Census field representatives that "all of the rooms or units at this building [were] used ENTIRELY to house people experiencing homelessness" (U.S. Census Bureau 2013).

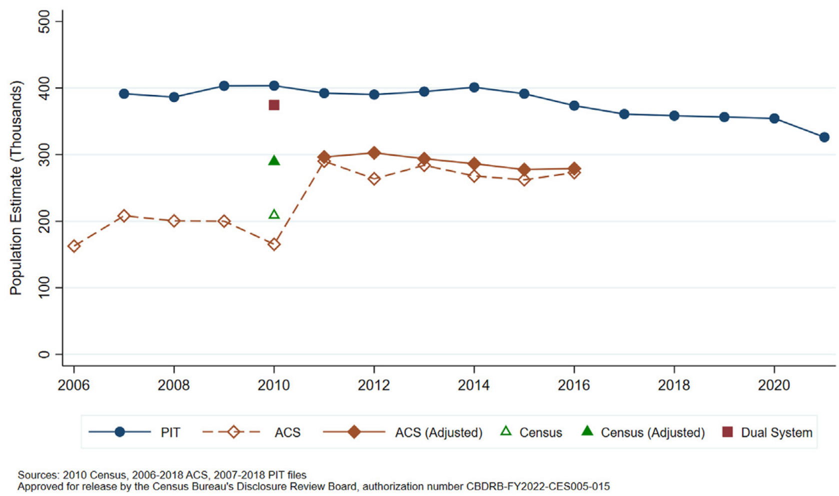


**Fig. 4.** Sheltered homeless population estimates in the PIT, ACS, and Census with definitional and weighting adjustments and dual system estimate.

benchmarks produced by the Census Bureau's Population Estimates Program (PEP). For the sheltered homeless, this scaling takes place within a broader class of group quarters types known as Other Non-Institutional (ONI) GQs, a category which also includes group homes, residential treatment centers for adults, workers' group quarters, and religious group quarters. The population benchmark for this group, however, is based on the most recent Census, and in the Census this category includes several additional types of group quarters that are outside the ACS's scope, namely unsheltered homeless locations, domestic violence shelters, and a few smaller categories. Fig. 3 provides a graphical representation of the ONI category and the various GQ types. The use of this broader population benchmark in constructing ACS weights means that the sheltered homeless population estimates are inflated to represent people who are not in the ACS's scope.

To correct this bias, we estimate the factor by which the ACS scales up the in-scope population each year by taking the ratio of the 2010 Census population in this ONI category that was in scope for both the Census and ACS to the population that was in scope only for the ACS. Dividing the ACS sheltered homeless estimate by this factor allows us to estimate our target, which is the sheltered homeless population size.

### 4.3. *Results from the comparison of aggregate estimates*

Fig. 4 presents sheltered homeless population estimates with definitional and weighting adjustments. With adjustments, the Census sheltered estimate rises from about 209,000 to more than 290,000, closing nearly half of the prior gap between the Census and PIT. Table 1 displays the year-by-year population estimates for each category of the PIT-only population. Domestic violence shelter occupants comprise the largest group, about 40,000 people each year. Voucher and non-shelter beds each contribute about 20,000 people each year.

Relative to the Census, the adjusted ACS estimates rise by a much smaller amount because the definitional adjustment, which increases the population estimate, is counteracted by the weighting bias correction. Table 2 displays the ACS in-scope and out-of-scope ONI populations in the 2010 Census and presents our estimate of the ACS scaling factor of about 1.32. In other words, we estimate that the ACS's person weights inflated the homeless population estimate by about 32 percent to represent people residing in domestic violence shelters, at unsheltered locations, and in other group quarters types outside the ACS's scope.

In the end, we are left with definition- and weighting-adjusted Census and ACS estimates that are about three-quarters of the PIT estimate in each year. We have reconciled about half of the initial gap between the Census and the PIT, representing about 80,000 people. In upcoming sections, we discuss potential explanations for the remaining gap between sources, such as shelter list completeness, ambiguity in the classification of certain facilities, and discrepancies arising from the timing of Census response.

## 5. Comparisons of Census and administrative shelter microdata

In this section, we compare Census and administrative shelter microdata to further explain the gap between the sheltered Census and PIT estimates. Specifically, we link HMIS data from Los Angeles and Houston to the 2010 Census using restricted linkage keys available on both sources. These links allow us to observe whether and in what housing status particular individuals from HMIS data were included in the Census. Because HMIS is a key data source for the PIT, this approach proves informative about the coverage and accuracy of both the Census and PIT.

### 5.1. *Assessing HMIS data quality*

We begin by assessing the quality of HMIS data with the goal of understanding how accurately these data represent those in shelters at a point in time. Accurate shelter entry and exit dates are critical to this section's analyses because they allow us to identify people who were in HMIS shelters during the Census. Fig. 5 displays the average daily shelter occupancy for Los Angeles from January 2009 to December 2013 as implied by HMIS entry and exit dates. We also indicate the number of HMIS beds available (shelter capacity) as indicated by the city's housing inventory count. In Los Angeles, capacity increases each winter as part of the city's Winter Shelter Program, which runs from December 1 to March 15. We extrapolate linearly from one year's point-in-time bed inventory to the next.

Several patterns in the Los Angeles data suggest errors in the exit dates recorded in HMIS in 2009–2011. First, we observe implausibly large increases in occupancy during these years' winter months, leading occupancy to far exceed capacity. We also observe precipitous drops on a handful of days, including March 31 of 2009 and 2010 and June 15 of 2011, suggesting that HMIS administrators conducted a purge of open shelter spells on those dates.[5] Analyses of shelter entry rates and hazard rates for shelter exit suggest that the above-described patterns are driven by incorrect exit dates, not incorrect entry dates.[6]

Fig. 6 displays daily occupancy and capacity in Houston HMIS data for 2009–2013. Unlike in Los Angeles data, we do not observe precipi-

[5] Los Angeles' Winter Shelter Program ends on March 15, so large drops on this day, but not other days, are consistent with the closing of seasonal shelters.

[6] Table A4 in the appendix display HMIS shelter entry rates (as a share of the 2010 Los Angeles population) by month for 2009-2013 and HMIS shelter exit hazard rates (i.e. the probability of exiting a shelter in a given month conditional

**Table 1**
Homeless Population Estimates.

| | Unadjusted estimates | | | PIT-only population estimates | | | | Adjusted estimates | | Dual system |
|---|---|---|---|---|---|---|---|---|---|---|
| Year | PIT (1) | Census (2) | ACS (3) | Safe Haven (4) | Domestic Violence (5) | Voucher-Based (6) | Non-Shelters (7) | Census (8) | ACS (9) | (10) |
| 2008 | 386,361 | – | 162,700 | – | 39,818 | 20,854 | 19,655 | – | – | – |
| 2009 | 403,308 | – | 208,200 | – | 39,156 | 20,902 | 19,655 | – | – | – |
| 2010 | 403,543 | 209,000 | 200,600 | 1,345 | 38,704 | 20,902 | 19,656 | 289,607 | – | 374,500 |
| 2011 | 392,316 | – | 200,200 | 1,898 | 37,127 | 21,757 | 16,041 | – | 296,354 | – |
| 2012 | 390,155 | – | 165,400 | 1,991 | 36,439 | 44,780 | 19,775 | – | 302,606 | – |
| 2013 | 394,698 | – | 290,000 | 2,025 | 35,431 | 20,602 | 20,797 | – | 293,767 | – |
| 2014 | 401,051 | – | 263,700 | 2,014 | 35,118 | 22,540 | 23,787 | – | 286,260 | – |
| 2015 | 391,440 | – | 283,900 | 1,861 | 34,483 | 20,202 | 22,387 | – | 277,495 | – |
| 2016 | 373,571 | – | 267,900 | 1,686 | 34,475 | 15,551 | 20,661 | – | 278,959 | – |
| 2017 | 360,867 | – | 262,300 | 1,463 | 34,241 | 14,277 | 27,729 | – | – | – |
| 2018 | 358,363 | – | 272,900 | 1,947 | 34,292 | 16,428 | 11,430 | – | – | – |
| 2019 | 356,422 | – | – | 1,933 | 34,469 | 12,636 | 14,494 | – | – | – |

**Source:** 2008–2019 Official PIT Files, 2008–2019 HIC Files, 2010 Census, 2008–2019 ACS.
**Note:** Table displays each year's PIT count as well as the number of people identified as being in Safe Haven beds by the official PIT files. Counts in domestic violence, voucher-based, and non-shelter beds are calculated by summing the PIT counts associated with people in each of these types of facilities in the HIC files. For some CoCs in some years, the HIC files lack PIT counts. In these cases, we impute the share of that CoC's PIT count in these types of beds using that CoC's share in the first subsequent year for which data is available. Adjusted Census estimate is calculated by adding PIT-only population estimates to Census total. Adjusted ACS estimate is obtained by adding PIT-only population estimates and then scaling down by the ACS scaling factor to correct weighting bias. Dual system estimate is obtained using methods described in Section 7 of the text. The estimate reported here is the midpoint of the range of estimates in that section.

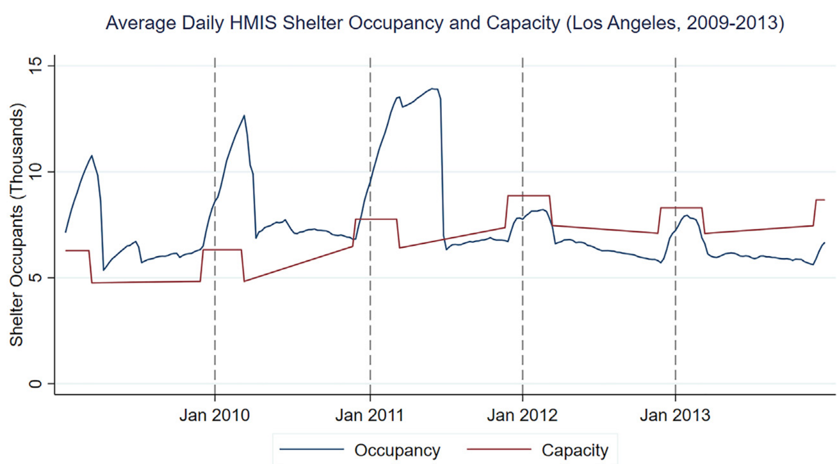


**Fig. 5.** Los Angeles HMIS Data Quality.

tous drops on specific dates or occupancy that exceeds capacity. We do not rule out the possibility of errors in recorded entry and exit dates in Houston, but we do observe that such errors, if they do exist, appear to arise in a less apparent and systematic fashion than in Los Angeles.

Several other pieces of evidence point to errors in the entry and exit dates recorded in HMIS. During the 2004–2014 period, we find that 21.4 percent of individuals have at least one instance of two or more overlapping emergency and transitional shelter spells, implying an erroneous entry or exit date for at least one of the spells. Moreover, using methods described in the next section, we estimate that 2.3–2.5 percent of people indicated by Los Angeles HMIS data as being in a shelter on April 1, 2010 were counted by the Census in local jails or state prisons on that day, and we take the Census status in these cases to be more reliable.[7,8]

### 5.2. Linking HMIS data to the Census

We link HMIS data to the 2010 Census using Protected Identification Keys (PIKs). The U.S. Census Bureau's Person Identification Validation System assigns PIKs to individuals who appear in survey or administrative data by searching for a matching record by Social Security Number (if available), name, date of birth, sex, and address in a reference file derived from SSA records and augmented with Individual Taxpayer Iden-

on being in the shelter at the beginning of the month). We observe similar trends in HMIS entry rates by month across years. Shelter exit hazard rates by month, by contrast, differ substantially across years. In 2009-2011, the hazard rate for exit in January or February is very low relative to 2012-2013; in March 2009-2010 and June 2011, in contrast, it is very high relative to those same months 2012-2013. This table suggests that it is the distribution of exit dates, not entry dates, driving excessive occupancy in the winter months of earlier years.

[7] Official HMIS documentation also acknowledges the possibility of incorrect date reporting. The 2014 HMIS data guide notes that some providers may enter clients into HMIS once they are "accepted" into a program, but prior to placing them in a bed. It also states that HMIS administrators "often forget to enter an exit date in HMIS for a client leaving the program since there is no operational trigger to remind them to do so" (U.S. Department of Housing and Urban Development 2012). The guide further states that some CoCs have a policy of auto-exiting open shelter spells after 90 days.

[8] In Houston, in contrast, the Census records less than one percent of HMIS shelter users as being in state prisons and local jails on a date when HMIS data indicated they were in shelters.

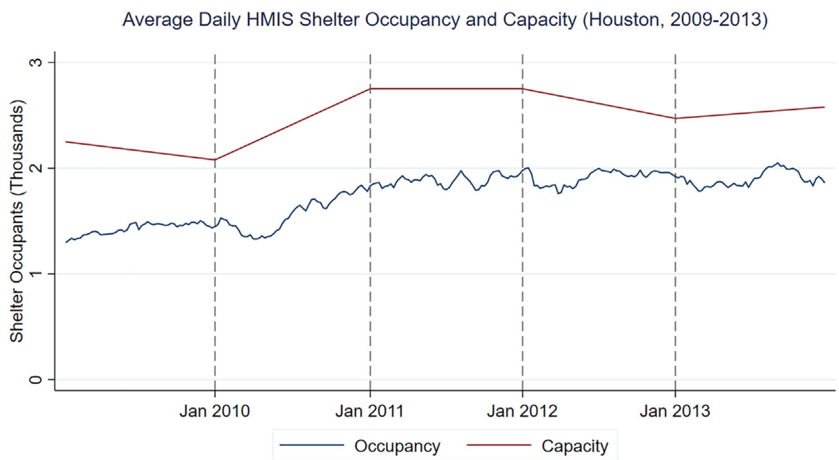


**Fig. 6.** Houston HMIS Data Quality.

**Table 2**
Population of Other Non-Institutional (ONI) Group Quarters (GQ) Types in the 2010 Census.

| | Population in 2010 Census |
|---|---|
| **A: Census and ACS Scope** | |
| Homeless Shelters | 210,036 |
| Group Homes | 307,129 |
| Residential Treatment Centers | 142,406 |
| Workers' Living Quarters | 169,107 |
| Religious Group Quarters (Est.)* | 75,684 |
| **Total** | 904,362 |
| **B: Census Scope Only** | |
| Soup Kitchens and Food Vans | 175,434 |
| TNSOLs | 37,502 |
| Maritime Vessels | 51,864 |
| Natural Disaster Shelters | 26 |
| Domestic Violence Shelters (Est.)* | 25,204 |
| **Total** | 290,030 |
| **ACS Scaling Factor** *(total of A plus B, divided by total of A)* | 1.321 |

**Source:** 2010 Census Service-Based Enumeration Assessment Report, 2010 Census Group Quarters Enumeration Assessment Report.
**Notes:** Table displays the population counts for various ONI GQ types in the 2010 Census, divided into those that are in-scope for both the Census and ACS and those that are in-scope for the Census only. *Indicates that these are estimates, not counts. The Census pools together religious GQs and domestic violence shelters in both public counts and restricted data. In the 2010 Census, this combined group had 100,888 people. We divide the group into a religious GQ estimate and a domestic violence estimate by assuming the ratio of the overall sheltered homeless population to the domestic violence population is the same in the PIT and the Census.

tification Numbers (ITINs) and other information (Layne and Wagner 2014).

Table 3 presents the share of records in our HMIS and Census datasets that are assigned a linkage key. Linkage rates are high for HMIS data because shelters frequently collect SSNs from service users. About 87.9 percent of Los Angeles HMIS shelter users and 95.5 percent of Houston HMIS shelter users in 2010 were assigned a linkage key. Census data do not contain SSNs, so linkage rates depend on the completeness and accuracy of personal information provided to enumerators, the uniqueness of this information, and the coverage of the reference file. Linkage rates for the Census data vary by enumeration site type. The linkage rates in the 2010 Census were 68.6 percent for the sheltered homeless, 42.4 percent for individuals at food vans, 41.8 percent for individuals counted using soup kitchens, and 17.2 percent for individuals at TNSOLs.

We account for incomplete linkage using inverse probability weights, which are estimated by obtaining the predicted probability of being assigned a linkage key in a probit model that accounts for individual characteristics recorded in the Census and HMIS data, including age, gender, race, Hispanic ethnicity. Ideally, we would like to weight our estimates by the inverse of the joint probability of being assigned a linkage key in both datasets, but we cannot directly estimate this target because our data do not allow us to differentiate between HMIS shelter users who do not appear in the Census because they truly were not counted and those who were in fact counted but were not assigned a linkage key. To address this challenge, we estimate bounds on the joint probability of being assigned a linkage key in both sources under the assumption that being assigned a linkage key in one source does not make an individual less likely to be assigned a key the other source. See Meyer et al. (2022) for more extensive discussion of this bounding methodology.

### 5.3. The coverage of HMIS shelter users in the Census

Table 4 displays estimated lower and upper bounds on the share of HMIS shelter users counted in various housing statuses in the Census in Los Angeles. We provide bounds on the coverage of all HMIS shelter users in the Census, as well as under three sets of refinements intended to drop individuals with incorrect exit dates. The first refinement drops individuals with an exit date of March 31, 2010, since the shelter occupancy patterns suggest a purge of open spells on that date. The second refinement drops individuals who were in shelters with names indicating participation in the city's winter shelter program, which ended on March 15, 2010. Refinement 3 further drops individuals with shelter entry dates prior to March 1, 2010, which is consistent with our understanding that entry dates recorded in HMIS are more reliable than exit dates. While this last refinement likely drops a large number of people who were truly in shelters on the Census date, we consider a comparison of results under the second and third refinements to be useful check for serious problems in HMIS entry and exit dates.

The share recorded as sheltered homeless in the Census increases with each refinement, suggesting that we have succeeded in better identifying people who were truly in shelters during the Census's homeless counting operation. Refinements 1 and 2 do not cause a large drop in the weighted count of people in Census shelters; most of the individuals dropped by these refinements are people who were counted as housed or had unknown status in the Census. Refinement 3, while allowing us to better identify a set of people who were truly in shelters, also causes the weighted count of people in shelters to drop substantially. We therefore suspect that most of the people dropped by refinements 1 and 2 were not in fact in HMIS shelters on March 30, 2010, whereas refinement 3 dropped many people who were truly in shelters on that date.

Under refinement 2, we estimate that 43–46 percent of HMIS shelter users were recorded by the Census in homeless shelters during the SBE. The range comes from the upper and lower bounds described in

**Table 3**
Linkage (PIK) Rates in Census and HMIS Data.

| | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **HMIS** | | | | | | | | | | | |
| Los Angeles[1] | 1.000 | 0.895 | 0.939 | 0.945 | 0.870 | 0.861 | 0.879 | 0.906 | 0.922 | 0.923 | 0.925 |
| Houston[2] | 0.800 | 0.949 | 0.979 | 0.967 | 0.955 | 0.956 | 0.955 | 0.961 | 0.962 | 0.965 | 0.965 |
| **Census** | | | | | | | | | | | |
| Shelter | | | | | | | 0.686 | | | | |
| Soup Kitchen | | | | | | | 0.418 | | | | |
| Food Van | | | | | | | 0.424 | | | | |
| TNSOL | | | | | | | 0.172 | | | | |

**Sources:** 2010 Decennial Census, 2004–2014 Los Angeles CoC HMIS Data, 2004–2014 Houston CoC HMIS Data.
**Notes:** Table reports the share of sheltered and unsheltered homeless individuals who are PIKed in the 2010 Census by GQ type. All results were approved for release by the Census Bureau, authorization number CBDRB-FY20-ERD002–004.

[1] Los Angeles Housing Management Information System (HMIS) data contains demographic and shelter use information for individuals who enrolled in emergency or transitional shelters in the Los Angeles CoC in 2004–2014. This CoC encompasses shelters in Los Angeles excluding Glendale, Long Beach, and Pasadena.

[2] Houston Housing Management Information System (HMIS) data contains demographic and shelter use information for individuals who enrolled in emergency or transitional shelters in the Houston CoC in years 2004–2015. This CoC encompasses shelters in Houston, Harris, Fort Bend, and Montgomery Counties.



**Table 4**
Coverage of Los Angeles and Houston HMIS Shelter Users in the 2010 Census.

| | Los Angeles | | | | | | | | Houston | |
|---|---|---|---|---|---|---|---|---|---|---|
| | All Records | | Refinement 1: Excluding 3/31 Exits | | Refinement 2: Excluding 3/31 Exits and Winter Shelter Program | | Refinement 3: Excluding 3/31 Exits, WSP, Entries Before 3/1 | | | |
| *Census Status* | *Lower Bound* | *Upper Bound* | *Lower Bound* | *Upper Bound* | *Lower Bound* | *Upper Bound* | *Lower Bound* | *Upper Bound* | *Lower Bound* | *Upper Bound* |
| Sheltered | 0.273 | 0.299 | 0.367 | 0.398 | 0.428 | 0.464 | 0.497 | 0.539 | 0.351 | 0.371 |
| Unsheltered | 0.106 | 0.119 | 0.105 | 0.116 | 0.085 | 0.092 | 0.156 | 0.170 | 0.035 | 0.037 |
| Other GQ | 0.077 | 0.088 | 0.068 | 0.076 | 0.071 | 0.079 | 0.047 | 0.054 | 0.151 | 0.160 |
| Housed | 0.267 | 0.292 | 0.236 | 0.253 | 0.236 | 0.252 | 0.193 | 0.205 | 0.218 | 0.226 |
| Status Unknown (not in Census) | 0.202 | 0.277 | 0.158 | 0.225 | 0.114 | 0.181 | 0.032 | 0.107 | 0.207 | 0.245 |
| Unweighted Total | 10,500 | | 7,000 | | 5,800 | | 1,300 | | 1,400 | |
| Share and PIKed in HMIS | 0.876 | | 0.886 | | 0.897 | | 0.923 | | 1.000 | |
| Share PIKed and in HMIS and Census | 0.522 | | 0.548 | | 0.577 | | 0.583 | | 0.536 | |
| Weighted Total | 10,420 | | 6,901 | | 5,738 | | 1,258 | | 1,480 | |

**Sources:** LA (CA-600, 2004–2014) HMIS administrative data, Houston (TX-700, 2004–2015) HMIS administrative data, 2010 Census.
**Notes:** Table displays the weighted share of individuals who were present in an emergency or transitional shelter in HMIS data on March 30, 2010, according to HMIS records, who appeared in the 2010 Census in various GQ types or as housed. Where exit dates were missing in HMIS data, we imputed an exit date based on the median stay length for users of that shelter type. Lower and upper bound weights calculated using methods described in the text. All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

**Table 5**
Coverage of HMIS Shelter Users in the 2010 Census by HMIS Program Type.

| | Los Angeles (Refinement 2) | | | | Houston | | | |
|---|---|---|---|---|---|---|---|---|
| | Emergency Shelters | | Transitional Housing | | Emergency Shelters | | Transitional Housing | |
| *Census Status* | *Lower* | *Upper* | *Lower* | *Upper* | *Lower* | *Upper* | *Lower* | *Upper* |
| Sheltered | 0.399 | 0.438 | 0.481 | 0.512 | 0.349 | 0.370 | 0.353 | 0.371 |
| Unsheltered | 0.108 | 0.118 | 0.041 | 0.045 | 0.062 | 0.065 | 0.020 | 0.021 |
| Other GQ | 0.070 | 0.080 | 0.072 | 0.076 | 0.011 | 0.012 | 0.225 | 0.237 |
| Housed | 0.172 | 0.185 | 0.351 | 0.372 | 0.143 | 0.151 | 0.260 | 0.269 |
| Status Unknown (not in Census) | 0.177 | 0.248 | −0.006 | 0.055 | 0.402 | 0.435 | 0.102 | 0.143 |
| Weighted Total | 3,697 | | 2,042 | | 533 | | 948 | |

**Sources:** LA (CA-600, 2004–2014) HMIS administrative data, 2010 Census.
**Notes:** Table displays the weighted share of individuals who were present in an emergency or transitional shelter in HMIS data on March 30, 2010, according to HMIS records, who appeared in the 2010 Census in various GQ types or as housed. Where exit dates were missing in HMIS data, we imputed an exit date based on the median stay length for users of that shelter type. Lower and upper bound weights calculated using methods described in the text. All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Section 5.2. About 8–9 percent were recorded in unsheltered statuses and about 7–8 percent were recorded in other group quarters facilities. About 24–25 percent were recorded as housed, and about 11–18 percent were not recorded in the Census and hence have unknown status. Our results indicate that about 82–89 percent of all people who were in HMIS shelters on that date were counted by the Census in some status.

The last two columns of Table 4 display bounds on the share of Houston HMIS shelter users who were recorded in various statuses in the Census. We do not make any refinements because we do not observe obvious, systematic errors in exit date reporting. We see that 35–37 percent of Houston HMIS shelter users were recorded by the Census in homeless shelters. About 4 percent were recorded as unsheltered homeless, 15–16 percent in other group quarters facilities, 22–23 percent in housing, and 21–25 percent with unknown status. These results are similar to those from Los Angeles, but with a few notable differences, including a smaller share recorded as sheltered or unsheltered homeless and a larger share with unknown status or in other group quarters. We also note that the weighted total number of HMIS shelter users was only about 1,500 in Houston, compared to about 5,700 under refinement 2 in Los Angeles.

#### *5.3.1. Explanations for status discrepancies between the Census and HMIS*

In this section, we explore potential reasons for discrepancies in individuals' statuses between HMIS data and the Census, including Census classification of certain HMIS shelters as housing units or other group quarters types, discrepancies arising from the timing of Census responses, and residual HMIS exit date errors.[9]

While Section 4 discussed straightforward differences in the definition of a shelter across sources, considerable definitional ambiguity remains. In particular, about 40 percent of HMIS shelter users on the PIT date in Los Angeles and Houston were in transitional shelters, which provide people experiencing homelessness a place to stay and supportive services for up to 24 months and typically require that residents possess a lease or occupancy agreement (U.S. Department of Housing and Urban Development 2018). Because these units provide longer-term and more stable tenure than emergency shelters, they are likely candidates for classification as housing units in the Census (Smith et al., 2012). Indeed, Table 5 shows that about 35–37 percent of people in transitional shelters in Los Angeles were recorded as housed in the Census, compared to just 17–19 percent of those in emergency shelters, with a similar pattern in Houston. Table 6 shows that about half of all HMIS shelter users in transitional housing in Los Angeles were in facilities where on average half of the residents were recorded in the Census as housed but none were recorded as sheltered homeless, a finding that further suggests that entire facilities were classified differently in the two sources.

We also find evidence that the Census classified some HMIS shelters as substance abuse treatment centers, group homes for adults with disabilities, or juvenile correctional facilities, discrepancies which may also have arisen from ambiguity in the definition of a homeless shelter. In Table 7, we see that of the 7–8 percent of Los Angeles HMIS shelter users recorded by the Census as being in other group quarters, about 43 percent were recorded in residential treatment centers for adults, which "provide treatment on-site in a highly structured live-in environment for the treatment of drug/alcohol abuse, mental illness, and emotional/behavioral disorders." The share of HMIS shelters users in this status rises when we refine our HMIS sample to exclude people with incorrect exit dates, suggesting that ambiguity in these facilities' classification, not incorrect exit dates, explains why shelter users are recorded in this status in the Census. Of the 15–16 percent of Houston HMIS shelter users in other group quarters in the Census, about one-fourth were recorded in group homes intended for adults, defined as "community-based group living arrangements that… provide room and board and services, including behavioral, psychological, or social programs." A key distinction between residential treatment centers and group homes is that the former emphasize substance abuse treatment, while the latter are targeted at people with physical and behavioral health conditions that require a supportive living environment. About 19 percent were recorded in a single correctional facility intended for juveniles, providing further evidence that entire facilities were classified differently in the two sources.

Another possible explanation for discrepancies in HMIS and Census statuses lies in the timing of Census responses from housing units. The SBE recorded individuals' housing status during a three-day window at the end of March, while the Census's housing unit questionnaire asked people to indicate their residence at the beginning of April. While a very small number of individuals may have transitioned from shelters to housing between these dates and hence been recorded as housed, a much larger number might have responded to the Census before entering a shelter or after exiting one during the long window of potential Census response. Census questionnaires were mailed to nearly all housing units on March 15, and by March 30, around half of these questionnaires had been received by the Census Bureau. The window of possible response also extended well beyond April 1, with about 20 percent of households responding during a non-response follow-up operation that began on May 1. Using the distribution of Census response dates and

[9] We also consider the possibility that erroneous linkages drive the observed discrepancies. Table A3 in the appendix displays the share of Los Angeles HMIS shelter users counted in the Census in various statuses who were found by the Census in California and in Los Angeles county. Nearly 90 percent of HMIS shelter users counted by the Census in unsheltered locations or other GQ types were found in Los Angeles, and about 97 percent were found in California. Among those counted by the Census as housed, 74.1 percent were in Los Angeles County and 84.9 percent were in California. These high geographic agreement rates do not suggest widespread erroneous linkages.

**Table 6**
HMIS Sheltered Individuals by Share Sheltered in Census and Census Status (Los Angeles).

| HMIS Shelter Type | Census recorded share in shelter | Bound | Share of People in Census Status | | | | Total People |
|---|---|---|---|---|---|---|---|
| | | | Shelter *(1)* | Housed *(2)* | Other Census Status *(3)* | Status Unknown *(4)* | *(5)* |
| Emergency | 0 | *Lower* | | 0.163 | 0.363 | 0.435 | 80 |
| | | *Upper* | | 0.146 | 0.419 | 0.475 | |
| Transitional | | *Lower* | | 0.537 | 0.113 | 0.336 | 850 |
| | | *Upper* | | 0.548 | 0.116 | 0.350 | |
| Emergency | 0 to 0.5 | *Lower* | 0.231 | 0.170 | 0.161 | 0.417 | 2,700 |
| | | *Upper* | 0.240 | 0.174 | 0.169 | 0.437 | |
| Transitional | | *Lower* | 0.363 | 0.152 | 0.136 | 0.335 | 350 |
| | | *Upper* | 0.369 | 0.153 | 0.143 | 0.349 | |
| Emergency | 0.5 to 1 | *Lower* | 0.811 | 0.016 | 0.047 | 0.115 | 550 |
| | | *Upper* | 0.819 | 0.016 | 0.051 | 0.125 | |
| Transitional | | *Lower* | 0.874 | 0.031 | 0.007 | 0.083 | 600 |
| | | *Upper* | 0.880 | 0.030 | 0.006 | 0.089 | |

**Sources:** 2010 Census, 2004–2014 Los Angeles HMIS data.
**Notes:** Sample is restricted to shelters with greater than ten occupants. Lower and upper bound weights calculated using methods described in the text. All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

**Table 7**
Distribution of Group Quarters Codes for HMIS Shelter Users Appearing in "Other GQ" Statuses in Census.

| GQ Code | Category | Los Angeles | | | Houston |
|---|---|---|---|---|---|
| | | All records | Refinement 1: Excluding 3/31 exits | Refinement 2: Excluding 3/31 exits and WSP | All records |
| 103 | State Prisons | 0.130 | 0.106 | 0.093 | – |
| 104 | Local Jails | 0.313 | 0.253 | 0.228 | – |
| 301 | Nursing Facilities | 0.063 | 0.073 | 0.089 | – |
| 203 | Correctional Facilities for Juveniles | – | – | – | 0.191 |
| 801 | Group Homes for Adults | – | – | – | 0.261 |
| 802 | Residential Treatment Centers for Adults | 0.278 | 0.407 | 0.430 | 0.513 |
| - | All Other GQ Codes | 0.217 | 0.161 | 0.160 | 0.035 |
| Overall share in Other GQs (midpoint of bounds) | | 0.083 | 0.072 | 0.075 | 0.155 |

**Sources:** L.A. and Houston HMIS administrative data, 2010 Census.
**Notes:** "HMIS shelter user" is defined as an individual who was in an HMIS shelter on March 30, 2010, according to HMIS administrative records. Dashed lines indicate categories that have been included in the "All Other GQ Codes" category due to the small number of observations in that category. All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

shelter entry and exit patterns, in conjunction with the distribution of Census response dates obtained from various Census press releases and official reports, we estimate that about 5.4 percent of HMIS shelter users might have been counted in housing before entering the shelter and 7.5 percent might have been counted as housed after leaving the shelter.[10] The timing of Census responses could therefore account for as much as half of the 24.4 percent of HMIS shelter users recorded by the Census as being housed.

Finally, some portion of status discrepancies can likely be attributed to remaining errors in HMIS exit dates. Table 7 provides some evidence of residual HMIS exit date errors. Under refinement 2, we observe that about 23 percent of the 7–8 percent of HMIS shelter users recorded by Census in other group quarters were found in state prisons and local jails. Because the Census enumeration in prisons and jails relied primarily on administrative records which are likely highly accurate, we interpret this as evidence of incorrect dates in HMIS data.

*5.3.2. Caveat on the calculation of housing status probabilities*

As a caveat, we note that the preceding analyses rely on the assumption that being assigned a linkage key is random conditional on the covariates in our inverse probability weighting model, meaning that the probability of being assigned a linkage key should be the same for a randomly chosen housed person as for a housed person who was recently homeless, given their covariates. If instead recently homeless individuals were less likely to be assigned a linkage key, then we would underweight those individuals, a tendency which could explain our observation that the share with unknown status (the residual category) decreases with each sample refinement, with each refinement disproportionately dropping underweighted rather than correctly weighted individuals. Recently homeless individuals who transition to housing could be difficult to link for various reasons, including the fact that they are less likely to be associated with their current address in the reference

[10] We use shelter entry and exit rates from 2012 because we are more confident in the accuracy of exit date reporting in this year than in 2010. Specifically, we considered the set of people who were in an HMIS shelter on March 30, 2012 and then for each date between March 1 and March 30, we multiplied the share of this group that entered the shelter on that day by the share of households that had responded to the Census by that day in 2010 according to Census reports. Then for each date March 31 to April 30 (the date after which the Census's non-response follow-up operation wound down) we multiplied the share of this group that exited the shelter on that day by the share of households that responded to the Census on or after that date in 2010 according to Census reports. We then summed these shares across all dates to obtain an estimate of the share of HMIS Census users who would have responded to the Census before entering the shelter or after exiting it. This estimate assumes that those who entered or exited the shelter had the same probability of responding in date range as the broader population.

files used for linkage.[11] Recently homeless individuals may also have a tenuous attachment to their living situation, meaning that the person responding to the Census questionnaire might lack complete or accurate information for them. Because about 90 percent of all housed people in the Census were assigned linkage keys, compared to 68 percent of the sheltered homeless, this issue could cause us to understate the count of people in housing and other group quarters by up to one-third and to overstate the count of people in the residual category (those with unknown status).

### 5.4. Double counting of homeless individuals in the Census

In this section we assess the extent of what turns out to be frequent double counting of homeless individuals in the Census. Table 8 displays weighted counts of HMIS shelter users from Los Angeles and Houston whose linkage key appears more than once in various combinations of Census statuses.[12] In Los Angeles, about 800–1,000 people, or 14–17 percent of the 5,800 HMIS shelter users, were counted in multiple statuses in the Census, most frequently in two housed statuses or in one housed and one sheltered homeless status. In Houston, about 10–11 percent of HMIS shelter users had a duplicate record.

Table 9 examines double-counting among all individuals counted in homeless statuses in the 2010 Census more broadly. Specifically, this table shows the share of all people counted in homeless shelters and in each of the unsheltered statuses who have at least one housed or other group quarters record in addition to their homeless record, as indicated by the presence of additional records with that same linkage key. We estimate that about 21-24 percent of the sheltered homeless, 45-56 percent of those counted in soup kitchens and food vans, and 29-35 percent of those at outdoor locations had at least one housed record in addition to their homeless record. About 1-3 percent of homeless individuals were included on some other group quarters record in addition to their homeless record. Among those with other group quarters records, the most common facilities were group homes, treatment centers, state prisons, and local jails.[13]

To understand the reasons for double counting, we first explore the possibility of erroneous linkage.[14] Table 10 displays agreement rates for age, gender, race, Hispanic ethnicity, and county and state of residence among duplicate record pairs in the Census. Among record pairs where a given characteristic is non-imputed for both records, sex matches in about 94 percent of cases. Agreement rates were also high for age, race, Hispanic ethnicity, and state and county. Several other facts give us confidence that duplication does not reflect widespread erroneous linkages. For one, we observe high rates of duplication even for HMIS shel-

[11] It is not necessary that the address on a Census record match the reference file for that record to be assigned a linkage key. Having a matching address in the reference file helps, however, because the Census Bureau's PIKing software uses address to narrow the scope of potential matches in the reference file and avoid duplicate matches.

[12] In previous analyses, we de-duplicated these records giving preference to sheltered, unsheltered, other GQ, and housed statuses, in that order.

[13] Duplication is a non-trivial issue in the Census more broadly. The 2010 Census Coverage Measurement (CCM) study found that about 2.8 percent of all person records in the 2010 Census were likely duplicates. A report from the Department of Commerce's Office of the Inspector General (OIG) described a particularly high risk of duplication for homeless individuals, which they attribute to official guidance that instructed enumerators to count homeless individuals even if they stated they had been previously counted at another service location, although the report also noted that this guidance was frequently ignored by enumerators (U.S. Department of Commerce 2011).

[14] We are unable to directly assess linkage quality because there is no single proxy for linkage error among records assigned a PIK by the Census Bureau's Personal Identification Verification System (PVS) (Abowd et al. 2020). (Layne et al. 2014) estimate aggregate false match rates for PVS, but these differ substantially depending on the nature of the input file and cannot be used to estimate probabilities of correct linkage at the record-to-record level.

**Table 8**
HMIS Shelter Users with Multiple Statuses in Census by Combination of Statuses.

| | Housed + | | | | Sheltered + | | | Unsheltered + | More than two | Other comb-ination | Shelter Users with Duplicate Records | Total Records |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Housed | Sheltered | Unsheltered | Other GQ | Sheltered | Unsheltered | Other GQ | Unsheltered | | | | |
| **Los Angeles** | | | | | | | | | | | | |
| **All records** | | | | | | | | | | | | |
| *Lower bound* | 247 | 376 | 131 | 144 | 19 | 37 | 86 | 16 | 91 | 24 | 1,172 | 10,500 |
| *Upper bound* | 289 | 479 | 188 | 199 | 20 | 57 | 134 | 19 | 139 | 48 | 1,570 | 10,500 |
| **Refinement 2** | | | | | | | | | | | | |
| *Lower bound* | 78 | 351 | 61 | 69 | 13 | 34 | 82 | – | 79 | 16 | 782 | 5,800 |
| *Upper bound* | 83 | 448 | 84 | 87 | 14 | 51 | 126 | – | 120 | 26 | 1,037 | 5,800 |
| **Houston** | | | | | | | | | | | | |
| **All records** | | | | | | | | | | | | |
| *Lower bound* | 22 | 67 | – | 19 | – | – | – | – | – | 31 | 139 | 1,400 |
| *Upper bound* | 23 | 71 | – | 20 | – | – | – | – | – | 32 | 147 | 1,400 |

**Sources:** Los Angeles HMIS data (2004–2014), Houston HMIS data (2004–2015).
**Note:** Table displays weighted counts of unique HMIS shelter users (as of 3/30/2010, without any restrictions) found in multiple statuses in the Census. In Los Angeles, in about 80% of cases the individuals' ages matched exactly, and about 90% of cases the individuals ages matched within five years. In about 95% of cases the individuals' sex matched. In about 92% of cases both individuals lived in California, and in about 88% of cases both individuals lived in L.A.county. All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization numbers CBDRB-FY2022-CES005-006 and CBDRB-FY2022-CES005-008.

**Table 9**
Homeless with duplicate housed or other GQ records in Census.

| Homeless type | All people | | Has at least one housed record | | | Has at least one other (non-homeless) GQ record | | |
|---|---|---|---|---|---|---|---|---|
| | Number of Records | Unique PIKs | Unique PIKs | Weighted population estimate | | Unique PIKs | Weighted population estimate | |
| | | | | Lower bound | Upper bound | | Lower bound | Upper bound |
| Shelter | 209,000 | 143,000 | 26,500 | 43,280 | 49,020 | 1,400 | 2,235 | 3,002 |
| Soup Kitchen | 162,000 | 67,000 | 29,000 | 72,670 | 84,800 | 1,200 | 2,924 | 4,078 |
| Food Van | 11,500 | 4,900 | 2,300 | 5,588 | 6,399 | 80 | 229 | 305 |
| TNSOL | 36,500 | 6,300 | 1,900 | 10,660 | 12,830 | 100 | 586 | 835 |
| | | | | Share of all records | | | Share of all records | |
| Shelter | | | | 0.207 | 0.235 | | 0.011 | 0.014 |
| Soup Kitchen | | | | 0.449 | 0.523 | | 0.018 | 0.025 |
| Food Van | | | | 0.486 | 0.556 | | 0.020 | 0.027 |
| TNSOL | | | | 0.292 | 0.352 | | 0.016 | 0.023 |

**Source:** 2010 Census.
**Notes:** Upper and lower bound weights estimated using methods described in the text. Among those with duplicate records in other GQ types, the most common GQ types for the sheltered homeless are state prisons (9.2%), local jails (23.1%), group homes (15.4%), and residential treatment centers (23.1%). The most common GQ types for the unsheltered homeless are state prisons (7.7%), local jails (23.1%), group homes (30.7%), and residential treatment centers (15.4%). All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization numbers CBDRB-FY2022-CES005-006 and CBDRB-FY2022-CES005-008.

**Table 10**
Agreement Rates for Characteristics of Duplicate Housed/Homeless Pairs in 2010 Census.

| | Imputed and Non-Imputed | | Non-Imputed Only | | Non-Imputed Only | | Non-Imputed Only | |
|---|---|---|---|---|---|---|---|---|
| | All records | | All records | | Same sex duplicates | | Different sex duplicates | |
| | Share | N | Share | N | Share | N | Share | N |
| Same sex | 0.937 | 59,500 | 0.939 | 57,000 | | 53,500 | | 3,500 |
| Age exactly the same | 0.709 | 59,500 | 0.775 | 53,000 | 0.819 | 47,500 | 0.099 | 3,100 |
| Age within one year | 0.756 | 59,500 | 0.811 | 53,000 | 0.855 | 47,500 | 0.120 | 3,100 |
| Age within five years | 0.867 | 59,500 | 0.903 | 53,000 | 0.942 | 47,500 | 0.299 | 3,100 |
| Same race | 0.812 | 59,500 | 0.851 | 51,000 | 0.862 | 46,000 | 0.670 | 2,800 |
| Same Hispanic status | 0.874 | 59,500 | 0.890 | 48,000 | 0.900 | 43,500 | 0.762 | 2,800 |
| Same state | 0.893 | 59,500 | 0.893 | 59,500 | 0.893 | 53,500 | 0.890 | 3,500 |
| Same county | 0.806 | 59,500 | 0.806 | 59,500 | 0.805 | 53,500 | 0.808 | 3,500 |

**Source:** 2010 Census.
**Note:** Table displays the share of duplicate housed/homeless pairs of records in Census for which the given characteristic is the same (or within a given interval) for both records. "Non-imputed" is defined here as having a flag indicating that a given characteristic was preserved "as reported" - i.e. not altered in any way (edited for consistency, allocated from hot deck). Sample includes only duplicate pairs where all characteristics are non-missing in both sources. All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-008.

ter users, and we are confident in the high quality of linkage keys assigned to these individuals because their records contain social security numbers. Second, in other work we observe that the sheltered and unsheltered homeless individuals counted in the Census experience persistently low income and high rates of program receipt over the course of a decade, even relative to a comparison group of poor single adults, patterns that we might not expect to see if housed individuals' linkage keys were erroneously assigned to homeless individuals in the Census (Meyer et al., 2022).

Misclassification offers another potential explanation for double counting. It is possible that Census enumerators classified individuals observed in soup kitchens and while using food vans as homeless when in fact those individuals were housed but happened to be using homeless services. However, we maintain that the potential for misclassification is quite low for people who were sleeping in homeless shelters and those counted on the streets at TNSOLs, because these individuals were classified based on where they spent the night. Frequent double counting even of people in these categories suggests that misclassification is not the predominant explanation.

Double counting might also occur if homeless individuals were included on the Census form of a housed family member or acquaintance with whom they occasionally resided. As discussed in previous sections, many people likely transitioned between homelessness and housing during the long window of Census response, a finding that could explain some double counting. Moreover, because the 2010 Census questionnaire instructed respondents to count all people "who live and sleep here most of the time," some homeless individuals may have been counted at the residence of a relative or acquaintance where they sometimes reside.

We explore this possibility in Table 11, which indicates the household characteristics of homeless individuals who are also included on a housed record. We see that about 19 percent of the sheltered homeless with a duplicate housed record are the only person residing in that housing unit, while the share ranges from 12–27 percent for the unsheltered depending on whether they were counted in a soup kitchen, food van, or TNSOL. The majority of homeless individuals with a duplicate housed record live with family. We also see that while the majority of those with a housed record appear on that record as the household head, a substantial share also appear as the child (typically the adult child) of the household head. Thus we see that in most cases, homeless individuals with duplicate housed records are not living alone and are in fact frequently living with family members. This pattern suggests that much of the observed double counting arises from these individuals' inclusion on the Census form of a family member or acquaintance.

**Table 11**
Household Characteristics of Homeless Individuals with a Duplicate Housed Record in 2010 Census.

| Homeless Record Type | Relationship to household head | | | | | Household type | | | N |
|---|---|---|---|---|---|---|---|---|---|
| | Household head | Spouse or partner | Child (adult or minor) | Other relative | Other nonrelative | Lives alone | Lives with family | Lives with non-family | |
| Shelter | 0.382 | 0.095 | 0.318 | 0.125 | 0.080 | 0.185 | 0.728 | 0.087 | 26,500 |
| Soup Kitchen | 0.516 | 0.120 | 0.183 | 0.100 | 0.081 | 0.268 | 0.616 | 0.117 | 29,000 |
| Food Van | 0.483 | 0.161 | 0.197 | 0.096 | 0.063 | 0.198 | 0.713 | 0.089 | 2,300 |
| TNSOL | 0.386 | 0.146 | 0.271 | 0.111 | 0.086 | 0.115 | 0.790 | 0.095 | 1,900 |

**Source:** 2010 Census.
**Note:** Sample includes all homeless individuals from 2010 Census with a single duplicate housed record. All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-008.

## 6. Dual system estimate of the sheltered homeless population

In this section, we use dual system estimation, a statistical technique widely employed in demography and other fields, to calculate a reliable estimate of the sheltered homeless population under certain assumptions. The U.S. Census Bureau has used dual system techniques to estimate under coverage in Decennial Censuses since 1980.[15] The first system consists of people enumerated in the Decennial Census and the second is an independent post-enumeration sample of the U.S. population. The share of people in the post-enumeration sample who were also found in the Census provides an estimate of the Census's coverage rate. Multiplying the Census count by the inverse of this share gives a consistent estimate of the true U.S. population under assumptions we discuss below (Wolter 1986).

In our context, the first system consists of those included in the Census's sheltered homeless count. The second consists of people who were in HMIS shelters on the day of the Census count in Los Angeles and Houston. The share of people in HMIS shelters on the day of the Census count who were found by the Census gives an estimate of the share of the true sheltered homeless found by Census. Multiplying the Census sheltered homeless count by the inverse of this share, we obtain an estimate of the true sheltered homeless population which is consistent if the number of individuals found in both samples is large and certain assumptions are met.

As an equation,

$$\begin{aligned}&Sheltered\ Homeless\ Estimate\\&= Census\ sheltered\ homeless\ count* \\&\times \frac{HMIS\ sheltered\ homeless\ count}{Count\ of\ HMIS\ sheltered\ homeless\ also\ found\ by\ Census\ in\ HMIS\ shelter}.\end{aligned} \tag{1}$$

We define the true sheltered homeless population to be people who were residing in facilities that align with the HMIS definition of a homeless shelter on the date of the Census count, recognizing that this excludes domestic violence shelters. We correct this definitional inconsistency at a later stage by adding an estimate of the population in domestic violence shelters to the estimate obtained from Eq. (1).

The Census definition of sheltered homelessness differs somewhat from the HMIS definition as well. In particular, the Census excludes from its sheltered homeless count those in voucher-funded hotel, motel and non-shelter beds and those in other facilities that HMIS classifies as shelters but Census classifies as housing or other group quarters, as suggested by the analyses in Section 5. We use (1) to account for these differences. Such individuals are appropriately included in the numerator of the ratio but not in the denominator. Eq. (1) also accounts for the extent to which the Census missed individuals in HMIS facilities that the Census defined as shelters.

We draw on results from our linked microdata comparisons in Section 5 to estimate the ratio on the right-hand side of (1). A complication in applying this framework is that errors in HMIS tend to prolong individuals' enrollments past their true exit dates. While we excluded some of these errors that were more easily identified in Section 5, other errors remain. For example, we believe all those found by the Census in jail or prison but recorded by HMIS as being in a shelter to be exit date errors. Such cases should be excluded from the numerator of the ratio on the right hand side of (1) because they were not in an HMIS shelter on the Census date. We must therefore estimate the number of HMIS observations that are from the time-period outside that of the Census homeless counting operation.

To do so, we estimate the share of those recorded erroneously in HMIS that is consistent with the count found in jail or prison by taking the share of HMIS shelter users found in jails or prisons in the Census and scaling it up by the inverse of the share of those leaving HMIS facilities that end up in jail or prison. We obtain an estimate of this latter ratio

[15] This approach is adapted from a method called "mark and recapture" often used in ecology to estimate the size of animal populations (McCallum 2000).

using the Census statuses of the sample of those who we identified as having date errors in Section 5, a group that we call HMIS shelter exiters. As an equation,

$$\frac{Count\ recorded\ erroneously\ in\ HMIS\ shelter}{HMIS\ sheltered\ homeless\ count} = \frac{Count\ found\ in\ jail\ or\ prison}{HMIS\ sheltered\ homeless\ count} * \times \frac{Count\ of\ HMIS\ shelter\ exiters}{Count\ of\ HMIS\ shelter\ exiters\ in\ jail\ or\ prison}. \quad (2)$$

As a final step, we must also estimate the count of people recorded erroneously in HMIS that were counted by the Census in non-HMIS homeless shelters. These are people who exited an HMIS shelter prior to the Census date but then entered a non-HMIS shelter and were counted there by the Census. This count, which is a subset of the overall count recorded erroneously in HMIS shelters that we subtracted from the numerator in (1), should also be excluded from the ratio's denominator because these individuals were not in HMIS shelters on the Census date. To estimate it, we take the share of those leaving HMIS facilities that ended up in non-HMIS shelters[16] and multiply this by the estimated count recorded erroneously in an HMIS shelter obtained using Eq. (2). We perform analogous calculations for the share that ended up in housing units, other group quarters, and unsheltered statuses and use these estimates to obtain estimates of the counts in these statuses after correcting exit date errors.

Table 12 displays counts and shares of the pooled Los Angeles and Houston samples in each Census status. We also indicate the share of HMIS exiters in each status and the HMIS sheltered homeless in each status after the date corrections described in this section. Applying counts in this table to Eq. (2), we estimate that about 36–38 percent of the HMIS sheltered homeless were erroneously recorded in an HMIS shelter due to incorrect dates. Scaling down the HMIS sheltered homeless count by 36–38 percent and assuming that these individuals are distributed across statuses in the Census according to the distribution of HMIS exiters' statuses, we obtain a date-corrected estimate of the share of the HMIS sheltered homeless in each status in column (4) of the table.

In summary, we estimate that about 60.8–63.8 percent of HMIS shelter users were found by the Census in shelters. Multiplying the inverse of this share by the Census sheltered homeless estimate of 209,000 as in Eq. (1), we obtain a non-domestic violence sheltered homeless estimate of about 328,000–343,000 people. To compare this estimate to the PIT, we add the approximately 39,000 people in domestic violence shelters to obtain a sheltered homeless population estimate of 367,000–382,000 people, or about 90–95 percent of the 2010 PIT count of about 403,500.

### 6.1. Assumptions of this methodology and caveats

Zhang (2019) formulates the assumptions of the dual system estimator in a setting where the researcher has access to population data from a population dataset (in our case, the Census sheltered homeless count), which is treated as fixed, and a population coverage survey (the HMIS data), which is treated as random.[17] Applying these assumptions to our setting, the dual system estimator from Eq. (1) will provide a consistent estimate of the true sheltered homeless population if four conditions are met. First, there must be no duplicated records or erroneous enumerations in either the HMIS or the Census homeless count. Second, the matched records between the HMIS and Census counts must be identified without errors. Third, the average HMIS capture probability for people in our Census dataset should be equal to the average HMIS capture probability for sheltered homeless individuals not in our Census dataset. And fourth, captures in the HMIS must be uncorrelated with one another, aside from intra-cluster correlations, which are permitted.[18]

To address the first assumption, we deduplicate records using linkage keys in both the HMIS and Census data and adjust for apparent exit date errors in HMIS to eliminate erroneous enumerations. After taking these steps, we are confident that the first assumption is reasonably close to satisfied. The second assumption relies on PIK-based linking being accurate which we believe to be a good approximation to the truth. Our inverse probability weights and bounding exercise address account for non-linkage. The fourth assumption, while difficult to test, strikes us as plausible because it allows for intra-cluster correlations (e.g. people residing in the same shelter may have correlated probabilities of inclusion in the Census).

The third assumption requires further discussion. For this assumption to hold, the average probability of inclusion in Los Angeles and Houston HMIS shelters among those in the Census sheltered homeless count must be equal to the average inclusion probability of all sheltered homeless individuals in the country. In 2010, the Los Angeles CoC estimated that about 40 percent of shelter users were in HMIS-tracked beds. Using the linked microdata, we estimate that 36–39 percent of the Los Angeles Census sheltered homeless were enrolled in HMIS shelters.[19] The similarity of these shares provides support for the third assumption in Los Angeles. Without additional HMIS data, however, we are unable to test this assumption for the U.S. sheltered homeless population more broadly.[20] This remains a caveat on our findings and a potential question for future work linking other localities' HMIS data to the Census.

## 7. Discussion

### 7.1. The size of the U.S. homeless population

A key goal of this paper was to triangulate homeless population estimates across available sources to improve our understanding of the U.S. homeless population size. We did so by comparing aggregate estimates as well as linked microdata and by using dual system methods to obtain a new estimate of the sheltered homeless population that is reliable under plausible assumptions. In this section, we discuss those findings' implications for the size of the U.S. homeless population. We also consider potential sources of bias in the Census and the PIT relative to the true homeless population. We discuss how these biases could affect aggregate comparisons and how they might explain differences between the PIT count and Census's sheltered homeless estimates and the dual system estimate.

#### 7.1.1. Unsheltered homeless population size

The 2010 PIT's unsheltered population estimate of 235,000 was similar to the Census's estimate of 210,000 people. We take this aggregate similarity to be encouraging, especially because this is the first time the widely cited PIT estimate has been compared to an independent national estimate. Aggregate comparisons, however, could mask bias in

[16] These shelters are necessarily non-HMIS because these are the people who were not in HMIS shelters at the time of the SBE.

[17] By treating the administrative list as fixed, this approach circumvents the problem of modeling the population dataset's potentially complicated data generating process. This approach also allows people who are and are not included in the population dataset to differ systematically from one another. The decision to treat the population dataset as fixed simplifies the assumptions for consistency from the extensive list described in (Wolter 1986).

[18] In the most basic formulation of these conditions, the third assumption states that HMIS capture probabilities must be constant for all sheltered homeless individuals and the fourth assumption states that captures in the HMIS must be uncorrelated with one another. Zhang (2019) shows that these assumptions can be relaxed to the formulations described in this text while preserving the consistency of the dual system estimator.

[19] See table A7 in the Appendix A.

[20] In Houston, we estimate that about 21-22 percent of the Census sheltered homeless were enrolled in HMIS shelters, a share that is well below the CoC's estimate that 60 percent of beds were tracked through HMIS that year. However, this discrepancy appears to be due in part to some HMIS shelters' exclusion from our internal files and in part to incompleteness in the CoC's inventory of non-HMIS shelters from those years.

**Table 12**
Weighted Counts and Shares of HMIS Shelter Users by Census Status (Los Angeles and Houston Pooled).

| A: Weighted counts | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | All records | | All records minus first set of those with exit date errors | | First set of those with exit date errors | | Records with correct dates | |
| | (1) | | (2) | | (3) | | (4) | |
| | *Lower bound* | *Upper bound* | *Lower bound* | *Upper bound* | *Lower bound* | *Upper bound* | *Lower bound* | *Upper bound* |
| Sheltered | 3368 | 3660 | 2976 | 3212 | 392 | 448 | 2750 | 2966 |
| Unsheltered | 1157 | 1294 | 537 | 584 | 620 | 710 | 180 | 194 |
| Other GQ | | | | | | | | |
| Non-Jail and Prison | 673 | 749 | 500 | 542 | 173 | 207 | 400 | 428 |
| Jail and Prison | 357 | 408 | 131 | 145 | 227 | 264 | 0 | 0 |
| Housed | 3101 | 3373 | 1674 | 1778 | 1427 | 1595 | 852 | 902 |
| | *Upper bound* | *Lower bound* | *Upper bound* | *Lower bound* | *Upper bound* | *Lower bound* | *Upper bound* | *Lower bound* |
| Status unknown | 3243 | 2414 | 1400 | 957 | 1843 | 1457 | 338 | 157 |
| Total | 11,899 | 11,899 | 7218 | 7218 | 4681 | 4681 | 4521 | 4648 |



| B: Weighted shares | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | All records | | All records minus first set of those with exit date errors | | First set of those with exit date errors | | Records with correct dates | |
| | (1) | | (2) | | (3) | | (4) | |
| | *Lower bound* | *Upper bound* | *Lower bound* | *Upper bound* | *Lower bound* | *Upper bound* | *Lower bound* | *Upper bound* |
| Sheltered | 0.283 | 0.308 | 0.412 | 0.445 | 0.084 | 0.096 | 0.608 | 0.638 |
| Unsheltered | 0.097 | 0.109 | 0.074 | 0.081 | 0.132 | 0.152 | 0.040 | 0.042 |
| Other GQ | | | | | | | | |
| Non-Jail and Prison | 0.057 | 0.063 | 0.069 | 0.075 | 0.037 | 0.044 | 0.089 | 0.092 |
| Jail and Prison | 0.030 | 0.034 | 0.018 | 0.020 | 0.048 | 0.056 | 0.000 | 0.000 |
| Housed | 0.261 | 0.283 | 0.232 | 0.246 | 0.305 | 0.341 | 0.188 | 0.194 |
| | *Upper bound* | *Lower bound* | *Upper bound* | *Lower bound* | *Upper bound* | *Lower bound* | *Upper bound* | *Lower bound* |
| Status unknown | 0.273 | 0.203 | 0.194 | 0.133 | 0.394 | 0.311 | 0.075 | 0.034 |
| Total | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |

**Source:** 2010 Census, 2004–2014 Los Angeles HMIS datasets, 2004–2015 Houston HMIS datasets.
**Notes:** Table indicates weighted counts in Census statuses, calculated as the sum of weighted totals from Los Angeles and Houston HMIS datasets. Columns (2) indicate bounds on the sum of Houston and L.A. weighted totals under Refinement 2. Columns (3) indicate bounds on the difference between (1) and (2). Columns (4) scales down the weighted total from (2) by one minus estimated share counted erroneously in an HMIS shelter (share in jail or prison in Columns (2) times the inverse of the share in jail or prison in Columns (3)), and then distributes these deletions according to the distribution of statuses in Columns (3).

each source relative to the true population, and we are unable to estimate this population using dual system methods because we lack a second source of microdata on these individuals. To address this concern, we discuss potential sources of bias in the Census and PIT relative to the true unsheltered homeless population and how biases might affect their aggregate difference.

We can characterize the relationship between each source's estimate and the true unsheltered homeless population on the PIT date ($H_{True}$) with the following equations:

$$H_{True} = H_{PIT} + U_{PIT} - O_{PIT}$$

$$H_{True} = H_{Census} + U_{Census} - O_{Census} + S$$

where $H_j$ for $j \in \{PIT, Census\}$ is the unsheltered estimate in a source, $U_j$ and $O_j$ are counts of people who were undercounted (missed in the PIT or Census) and overcounted (double counted or misclassified as unsheltered), and $S$ is the seasonal difference in true population sizes (at the time of the PIT minus the Census).

Combining these expressions shows that the aggregate difference between the PIT and Census reflects the difference between each source's net error ($U_j - O_j$) and seasonal differences:

$$H_{PIT} = H_{Census} + \left(U_{Census} - O_{Census}\right) - \left(U_{PIT} - O_{PIT}\right) + S$$

We are interested in the magnitude of each source's net error because this indicates bias relative to the true unsheltered homeless population. An aggregate comparison does not allow us to estimate net error in each source, but it does tell us about the difference of net error. Results based on Meyer et al. (2022), which compares the ratio of Census to PIT counts across CoCs accounting for several measures of temperature and precipitation, suggest that $S$ is small, with reasonable estimates ranging from about −8 to 6 percent of the Census unsheltered count. We therefore emphasize sources of over and undercounting in this section.

Overcounting could arise in either source from the misclassification of housed or sheltered homeless people as unsheltered. Both the Census and the PIT obtain unsheltered estimates in part from counting people using homelessness services. While both sources' methodology documents instruct those doing the count to ask people's unsheltered status, it is possible that the chaotic nature of such locations made it impossible to correctly determine everyone's unsheltered status, leading to misclassification. However, such misclassification appears to be small in the Census. Only 2 percent of the Census unsheltered homeless in Houston and 4–5 percent of those in Los Angeles were enrolled in HMIS shelters on the SBE date, an occurrence that could reflect either misclassification or incorrect HMIS shelter exit dates. We do not have an estimate of misclassification in the PIT.

Overcounting could also arise due to double counting during both sources' multi-day counting operations. This is likely a minor source of bias in the Census because the Census's post processing algorithm deduplicated records within the universe of homeless records using personal information. Table 8 shows that it is very rare for someone to be counted multiple times in sheltered or unsheltered homeless statuses in the Census, although a caveat on this is that people who did not provide personal information cannot be deduplicated. CoCs, on the other hand, rarely collect personal information from unsheltered homeless individuals when conducting the PIT counts, so deduplication methods are much less sophisticated, typically consisting of simply asking whether people have already been counted (HUD 2014).

Overall, we suspect that double counting and misclassification are more important sources of bias in the PIT than in the Census because its counting operations often rely on volunteers with minimal training whose understanding of and fidelity to protocols may be limited. Moreover, CoCs apply for funding based on the outcome of the PIT count and hence may not be indifferent to their outcomes. If overcounting is more widespread in the PIT count than the Census, then this would explain some of the aggregate difference between sources.

We also consider potential bias from undercounting in each source. Because both the PIT and Census rely on finding people at service locations and on canvassing outdoor locations at night, both would tend to miss people who do not use services or choose to sleep in isolated or hidden locations, such as vehicles or abandoned buildings. This could lead to correlated undercounting in the sources that would net out in an aggregate comparison. We therefore expect that some amount of undercounting is present and that the magnitude may be similar in both sources, but we are unable to estimate this bias using available data.

In summary, we expect both sources' unsheltered estimates to be biased to some extent by under and overcounting, but these biases are difficult to estimate. We suspect that greater duplication and misclassification in the PIT count could explain some of the aggregate differences between sources. Undercounting may be important in both sources and could net out in aggregate comparisons, but without estimates of overcounting we cannot determine the sign or magnitude of net bias in each source's estimate relative to the true population. Given the substantial difficulties of counting this population and methodological differences between the PIT and Census, the fact that both arrive at similar results provides encouraging evidence that both offer reasonable estimates of the unsheltered population size.

#### *7.1.2. Sheltered homeless population size*

Prior to adjustments, the Census's sheltered homeless estimate of about 209,000 people fell far short of the 2010 PIT estimate of about 405,000. The ACS estimate was about half of the PIT in 2006–2010 and about three-quarters of the PIT after 2010. However, we reconciled much of this initial discrepancy by accounting for straightforward definitional differences across sources and bias arising from the ACS weighting methodology. Specifically, we found that the Census SBE's exclusion of domestic violence shelters, voucher-funded hotel and motel beds, and beds in non-shelter facilities explained about half of the initial gap between the 2010 PIT and Census. People in these groups were counted in the Census but not classified as homeless. We also adjusted the ACS upwards to reconcile definitional differences, but then scaled down estimates by about 30 percent to correct bias arising from the ACS's weighting methodology. These straightforward definitional and weighting adjustments closed about half of the initial gap between the 2010 PIT count and the Census, leaving us with a definitionally-adjusted Census estimate of about 289,500.

Using the dual system methodology described in Section 6, we obtained a new sheltered homeless estimate of 367,000–382,000 people, or about 5–10 percent lower than the 2010 PIT estimate and about 27–32 percent larger than the adjusted Census count. Because this estimate did not make assumptions on the completeness of the PIT or Census, we maintain that this is a reasonable estimate of the sheltered homeless population size. Our findings in Sections 5 and 6 suggest that much of the gap between the adjusted Census count and the dual system estimate reflects ambiguity in the definition of a homeless shelter leading to a different classification of these structures by the Census rather than their omission. This explanation is consistent with findings in Meyer et al. (2022), which compares CoC-level sheltered population estimates in the 2010 Census and PIT and finds that on average, a given CoC has about three-quarters as many unique shelter addresses underlying its Census estimate as its PIT estimate and differences in facility count explain much of the gap between the aggregate adjusted Census and PIT sheltered population estimates. We next turn to a discussion of potential sources of bias in the sheltered PIT and Census and discuss how bias might explain differences between those sources' estimates and the dual system estimate.

The PIT could overstate the sheltered homeless population due to its reliance on HMIS data, which in the years around the 2010 Census tended to overstate the number of people enrolled in a shelter at a point in time. This issue would be a major concern if CoCs simply extrapolated from HMIS data to obtain their sheltered estimates, but in practice HUD instructs CoCs to implement a series of quality checks before using these data in their counts (HUD 2012). For example, in a 2010 report to HUD, the Los Angeles CoC stated that they compared shelters' capacity and occupancy and corrected counts where necessary when generating their sheltered PIT estimate. Such checks may not have caught all date errors, however, potentially leading to overcounting that could explain why the dual system estimate is lower than the PIT.

Double counting, on the other hand, is less of a concern in sheltered estimates because both the PIT and Census deduplicated sheltered homeless counts using personal information, including name and date of birth in the case of the Census and SSNs recorded in HMIS in the case of the PIT. However, in both sources, incomplete collection of personal information prevents comprehensive deduplication, and some double counting could remain.

Undercounting could have occurred in either source due to shelter list incompleteness. We have also seen that the Census appeared to classify many HMIS facilities as housing or other types of group quarters rather than as homeless shelters, a fact that would lead the Census estimate to understate the population relative to our target definition, which is based on the HMIS and PIT definition. Although we accounted for straightforward definitional differences in aggregate comparisons, microdata comparisons suggest that more subtle differences in classification remain. The combination of Census undercounting and residual classification differences may explain the difference between the Census and dual system estimates, with the latter estimate having corrected for both of these sources of undercounting.

### 7.2. Completeness and accuracy of available data on homelessness

The second major goal of this paper was to learn about the completeness and accuracy of available datasets on the U.S. homeless population, particularly the 2010 Census. Overall, we found that the coverage of sheltered homeless individuals in the 2010 Census was surprisingly good. Our dual system estimate implied that about 93–97 percent of people who were in HMIS shelters on the night of the Census's homeless counting operation were included in the Census in some status. Potential bias from the underweighting of people found as housed or in other group quarters, as described in Section 5, means that the true share could be even higher. About 61–64 percent were found by the Census in shelters, 19 percent in housing units, and 9 percent in other types of group quarters. The last 4 percent appear to have been misclassified as unsheltered.

As documented in Section 5, it appears that many of the HMIS shelter users not found in shelters in the Census were in facilities that the Census classified as housing or other types of group quarters. This pattern in part reflects the straightforward definitional differences identified in Section 4. In many cases, however, this pattern also appears to reflect more subtle distinctions in how HMIS and the Census define homeless shelters. For example, we found evidence that the Census classified many HMIS transitional shelters as housing, likely because the people residing there had fairly long-term and stable occupancy agreements. The Census also appears to have classified some HMIS facilities not as homeless shelters but as group homes for adults or residential treatment centers for substance abuse, meaning that those facilities' administrators chose that designation when asked by Census advance visit teams which group quarters type best described their facility. This finding highlights the lack of consensus about what types of facilities constitute a homeless shelter. This ambiguity, in turn, appears to matter substantially for estimates of the sheltered homeless population size.

Unexpectedly, our analyses also uncovered a pattern of frequent double counting of homeless individuals in the Census, often in a combination of housed and homeless statuses. Additional analyses suggested that most double counting arose because people transitioned from being housed to homeless around the time of the 2010 Census or because they were included on the Census form of a family member or acquaintance with whom they sometimes resided. Incorrect linkage and misclassification of housed individuals as homeless may in part explain double counting but do not appear to be its primary causes. These findings illustrate the fluidity of homeless individuals' living situations between housed and homeless statuses.

Finally, our analyses revealed important issues with the quality of exit dates recorded in HMIS data, which are widely used by both program administrators and homelessness researchers. In 2009–2011 in Los Angeles, shelter occupancy, as indicated by HMIS entry and exit dates, far exceeded capacity in winter months and dropped precipitously on a handful of dates, suggesting a purge of open shelter spells. We also found frequent instances of overlapping shelter spells, and we obtained further evidence of errors in the form of individuals who were found in state prisons and local jails during the 2010 Census despite being enrolled in the shelter according to HMIS data. These findings recommend caution for researchers using these data to identify people in shelters at a point in time or to analyze temporal patterns of shelter usage.

## Conclusions

Our work suggests that on any given night, there are about 600,000 people experiencing homelessness in the U.S. and that about one-third are sleeping on the streets and the rest in shelters. We estimate that the 2010 sheltered homeless population was about 367,000–382,000, a range that is slightly lower than HUD's widely cited point-in-time estimate and much larger than the Census's sheltered homeless count, with the latter fact due largely to differences in how HUD and Census defined a homeless shelter. Our work suggests that the Census estimate of 210,000 and the PIT estimate of 235,000 provide a reasonable range for the unsheltered homeless population size, although we acknowledge the possibility of under or over counting in each source. The dual system methods used in this paper may prove useful to other researchers looking to estimate the unsheltered homeless population size, although doing so requires a set of linkable data on the unsheltered population that satisfies the assumptions of this method. Taken together, the findings in this paper lend new credibility to aggregate PIT estimates that had not previously been validated against independent estimates. At the same time, they highlight the fact that there is considerable ambiguity about what types of facilities constitute a homeless shelter and that population estimates are very sensitive to these ambiguities.

Our work also suggests that most homeless individuals were included in the Census, although they were oftentimes counted as housed or in other types of group quarters. Many were counted twice, reflecting frequent transitions between housing status even in a dataset designed to convey a static picture of the U.S. population. This finding has implications for the coverage of homeless individuals in household surveys other than the ACS, like the Current Population Survey (CPS) and Survey of Income and Program Participation (SIPP), which are not intended to represent the homeless population. Given the frequency of double counting, we suspect that homeless individuals may in fact be included in surveyed households' responses more often than previously thought. These findings contribute to a larger emerging picture of the mobility and persistent material deprivation of the U.S. homeless population.

The Census and ACS hold tremendous promise for learning about homelessness. By establishing the broad coverage and reliability of the new data sources, our analyses lay the foundation for pathbreaking work using these data sources to learn about the demographic characteristics, income, safety net program participation, mortality, housing transitions, and migration patterns of those experiencing homelessness, work that promises to advance substantially our understanding of this difficult to study population.

## CRediT authorship contribution statement

**Bruce D. Meyer:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Angela Wyse:** Conceptualization, Methodology, Software, Data curation, Writing – original draft. **Kevin Corinth:** Conceptualization, Methodology.

## Appendix A

### A1. Comparison of sheltered homeless characteristics across sources

We also compare the characteristics of sheltered homeless individuals in the PIT, Census, and ACS to assess the extent to which they represent the same population. Table A1 reports the share under 18, gender/sex, race, and Hispanic ethnicity of sheltered individuals in the 2016 ACS and PIT.[21] The share belonging to various race categories and the share Hispanic are similar across the two data sources. The share female, however, is about 5 percentage points higher in the PIT (44.4 percent, compared to 39.4 percent in the ACS) and the share under age 18 is about 17 percentage points higher in the PIT (29.1 percent, compared to 12.2 percent in the ACS).

**Table A1**
Characteristics of Sheltered Homeless in PIT and ACS (2016).

| Source<br>*Includes Domestic Violence?* | ACS<br>*No* | PIT<br>*Yes* |
|---|---|---|
| **Age** | | |
| Under 18 | 0.122 | 0.291 |
| 18 and Older | 0.878 | 0.709 |
| **Gender/Sex*** | | |
| Male | 0.606 | 0.554 |
| Female | 0.394 | 0.444 |
| Other Gender | – | 0.002 |
| **Race** | | |
| White | 0.430 | 0.439 |
| Black | 0.454 | 0.451 |
| Asian | 0.018 | 0.009 |
| Am Ind/Pac Isl | 0.038 | 0.033 |
| Other Race (incl multiple) | 0.060 | 0.067 |
| **Hispanic Ethnicity** | | |
| Hispanic | 0.224 | 0.233 |
| Non-Hispanic | 0.776 | 0.767 |

**Sources:** ACS 2016 one-year estimates, 2016 PIT file.
**Notes:** ACS results approved for disclosure, CBDRB-FY20-ERD002–004. PIT and HMIS results obtained from public sources. *ACS collects data on sex. PIT collects data on gender, including transgender and gender non-conforming.

A back-of-the-envelope analysis suggests that the PIT's inclusion of domestic violence shelter residents could explain much of the gender discrepancy but little of the age discrepancy. For 2016, we estimate that about 9.2 percent of the sheltered PIT population consisted of people in domestic violence shelters. If we assume that all adults in domestic violence shelters were female and accompanied by one child on average, who was equally likely to be male or female, then removing domestic violence shelter occupants from the PIT would decrease the share female to 41.3 percent and decrease the share under 18 to 27.0 percent. Such an adjustment would therefore close most of the gap in the share female, but only a small portion of the gap in the share under 18.

Table A2 compares the share female and the share under 18 in the 2010 ACS to that in the Census. We observe that the share female is similar in these two sources, while the share under 18 is about 5 percentage points lower in the ACS than in the Census. This comparison once again suggests that the ACS may have missed some of those under 18. This finding suggests the need for caution in analyses studying the child homeless population using the ACS, but is reassuring for analyses that are limited to adults, such as studies of income and safety net program participation. We revisit this puzzle about differences in share of children across sources in Section A3.

### A2. Characteristics of recent HMIS shelter occupants missed by the Census

Los Angeles HMIS shelter users dropped by refinements 1 and 2 were disproportionately likely to have unknown status. The weighted count of people with unknown status fell from about 2500 prior to refinements to fewer than 1000 after refinements 1 and 2, where this weighted count is taken as the share of shelter users under a given refinement that fall into the residual category. In this section, we describe the characteristics of those individuals and discuss implications for the Census's coverage of the homeless and recently homeless population.

We know from HMIS shelter names that most of the people dropped in refinements 1 and 2 were participants in Los Angeles's Winter Shelter Program, which runs from December 1 to March 15 of each year. Unfortunately, because "status unknown" is a residual category, we do not know precisely which of the individuals dropped from the HMIS data fell into this category. We can, however, compare the overall characteristics of those who were kept and those who were dropped, as seen in Table A5. We observe that dropped individuals – those who were disproportionately likely to have unknown status – were older, more white, more Hispanic, and more male. They also had more frequent but shorter HMIS spells between 2004 and 2014.

One hypothesis is that these individuals were missed by the Census because they migrated to Mexico. We do indeed find that dropped individuals are more likely to be Hispanic (39 percent) than kept individuals (30 percent), but not overwhelmingly so. Another hypothesis is that these individuals may have transitioned to marginal living situations like couch-surfing, where they might have been left off the housing unit questionnaire submitted to Census. A third hypothesis that these individuals transitioned to unsheltered status. This hypothesis aligns with the Winter Shelter Program's primary purpose of shielding homeless individuals who would otherwise be unsheltered from the elements during the winter. Prior work has shown that unsheltered individuals tend to be older, more white, and more male that sheltered individuals, so these individuals' characteristics align with that profile (Meyer et al., 2022).

Taken together, the available evidence does not provide satisfactory resolution to the puzzle of why recent participants in Los Angeles's Winter Shelter Program were disproportionately likely to be missed by the Census. This group does, however, offer concrete evidence of a subset of recent shelter occupants who were missed by the Census.

### A3. Coverage of homeless children in linked HMIS-Census data

We also use the linked Census-HMIS data to revisit the puzzle identified in our aggregate comparisons section on the difference in the share of homeless individuals under age 18 in the PIT versus the ACS and Census. Table A6 displays the share of Los Angeles and Houston HMIS shelter users in various Census status disaggregated into those under 18 and those 18 and older. In contrast to our findings in Section 4, which suggested that children in the PIT were under-covered in the Census homeless enumeration, we see that about 48–52 percent of children in HMIS shelters were counted in homeless shelters in the Census, compared to 40–43 percent of adults. Children were also more likely to be counted as housed (30–32 percent) than adults (22–23 percent). We note that in 2010, HMIS data would likely not have included many facilities intended for unaccompanied youth because there was a separate system for tracking shelters intended for runaway and homeless youth prior to 2015. It is also possible that the Census classified some youth shelters as non-shelter facilities, as we found to be the case for some adult-oriented HMIS shelter. In Houston, we note that about 20 HMIS shelter users were counted in a single juvenile correctional facility in the Census, providing strong evidence of differential classification between sources in at least this instance. Table A1, Table A2, Table A3, Table A4, Table A5, Table A6, Table A7[22]

[21] The PIT did not report characteristics at this level of detail prior to 2015.

[22] This number is rounded per Census Bureau disclosure rules and has been reviewed for unauthorized disclosure of confidential information. The Census Bureau has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-008.

**Table A2**
Share Under 18 and Share Female of Sheltered Homeless in ACS, Census, and PIT.

| | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Under Age 18** | | | | | | | | | | | | | | |
| **ACS** | 0.178 | 0.189 | 0.159 | 0.131 | 0.153 | 0.135 | 0.104 | 0.133 | 0.158 | 0.128 | 0.122 | | | |
| **Census** | | | | | 0.202 | | | | | | | | | |
| **PIT** | | | | | | | | | | 0.292 | 0.291 | 0.286 | 0.282 | 0.273 |
| **Female** | | | | | | | | | | | | | | |
| **ACS** | 0.384 | 0.426 | 0.364 | 0.369 | 0.379 | 0.388 | 0.364 | 0.403 | 0.397 | 0.374 | 0.394 | | | |
| **Census** | | | | | 0.379 | | | | | | | | | |
| **PIT** | | | | | | | | | | 0.445 | 0.444 | 0.445 | 0.447 | 0.441 |

**Sources:** 2006–2016 ACS one-year estimates, 2010 Census, 2015–2019 PIT.
**Notes:** Table displays the share of sheltered homeless individuals in the 2006–2016 ACS, 2010 Census, and 2015–2019 PIT who fall into a given age or gender category. The ACS shares are weighted using survey weights prior to 2011. From 2011 onwards, we include only non-imputed ACS records, which are scaled up by a constant such that the new weighted count of non-imputed observations is equal to the old weighted sum of imputed and non-imputed records. All results were approved for release by the Census Bureau, authorization number CBDRB-FY20-ERD002–004.

**Table A3**
Share of HMIS Shelter Users in a Given County/State in the Census, by Housing Status in Census.

| Status in Census | County in Census | | State in Census | |
|---|---|---|---|---|
| | L.A. | Other | CA | Other |
| Sheltered | 0.956 | 0.044 | 0.978 | 0.022 |
| Unsheltered | 0.928 | 0.072 | 0.962 | 0.038 |
| Other GQ | 0.863 | 0.137 | 0.971 | 0.029 |
| Housed | 0.741 | 0.259 | 0.849 | 0.151 |

**Sources:** 2010 PIT, 2010 Census.
**Notes:** Table displays weighted share of HMIS shelter users who were in a given county or state in the Census, by housing status. Weight is calculated as the midpoint of the upper bound weight and the lower bound weight. All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

**Table A4**
Probability of L.A. HMIS Shelter Entry and Hazard Rate for Exit.

| *Entry Probability* | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| 2009 | 0.34 | 0.28 | 0.20 | 0.18 | 0.16 | 0.16 | 0.14 | 0.14 | 0.16 | 0.15 | 0.16 | 0.49 |
| 2010 | 0.40 | 0.30 | 0.28 | 0.31 | 0.27 | 0.21 | 0.19 | 0.18 | 0.19 | 0.18 | 0.17 | 0.57 |
| 2011 | 0.40 | 0.35 | 0.27 | 0.17 | 0.16 | 0.17 | 0.20 | 0.18 | 0.19 | 0.18 | 0.19 | 0.55 |
| 2012 | 0.40 | 0.33 | 0.29 | 0.21 | 0.19 | 0.16 | 0.16 | 0.16 | 0.13 | 0.14 | 0.16 | 0.54 |
| 2013 | 0.43 | 0.27 | 0.20 | 0.14 | 0.14 | 0.12 | 0.14 | 0.12 | 0.12 | 0.14 | 0.09 | 0.41 |
| *Hazard Rate for Exit* | | | | | | | | | | | | |
| Year | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
| 2009 | 0.08 | 0.08 | 0.72 | 0.12 | 0.11 | 0.30 | 0.13 | 0.14 | 0.18 | 0.14 | 0.14 | 0.18 |
| 2010 | 0.12 | 0.11 | 0.67 | 0.19 | 0.15 | 0.23 | 0.15 | 0.15 | 0.18 | 0.18 | 0.19 | 0.20 |
| 2011 | 0.15 | 0.10 | 0.18 | 0.08 | 0.07 | 0.71 | 0.17 | 0.19 | 0.18 | 0.21 | 0.21 | 0.29 |
| 2012 | 0.30 | 0.29 | 0.46 | 0.23 | 0.23 | 0.21 | 0.19 | 0.19 | 0.17 | 0.20 | 0.30 | 0.23 |
| 2013 | 0.29 | 0.31 | 0.38 | 0.17 | 0.19 | 0.18 | 0.17 | 0.17 | 0.16 | 0.19 | 0.17 | 0.21 |

**Sources:** L.A. HMIS data (2004–2014).
**Notes:** Table displays the probability of entering an L.A. HMIS shelter in a given month and year as a share of the 2010 Los Angeles population and the probability of exiting an HMIS shelter in a given month/year conditional on being in the shelter on the first day of the month. All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

**Table A5**
Characteristics of People Kept and Dropped in Refinements 1 and 2.

| | Kept | Dropped |
|---|---|---|
| Age at First Entry (Mean) | 35.60 | 40.52 |
| White (Share) | 0.39 | 0.52 |
| Black (Share) | 0.52 | 0.35 |
| Other Race (Share) | 0.09 | 0.13 |
| Hispanic (Share) | 0.30 | 0.39 |
| Female (Share) | 0.41 | 0.27 |
| Enrolled in Emergency Shelter (Share) | 0.61 | 0.99 |
| Number of Spells (2004–2014) (Mean) | 3.74 | 4.29 |
| Average Spell Length (Mean) | 216.70 | 75.35 |

**Sources:** LA (CA-600, 2004–2014) HMIS administrative data.
**Notes:** All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005–006.

**Table A6**
Coverage of HMIS Shelter Users in the 2010 Census by Child/Adult (L.A. and Houston Combined).

| | Children (Age < 18) | | Adults (Age 18+) | |
|---|---|---|---|---|
| *Census Status* | *Lower Bound* | *Upper Bound* | *Lower Bound* | *Upper Bound* |
| Sheltered | 0.482 | 0.524 | 0.403 | 0.434 |
| Unsheltered | 0.001 | 0.001 | 0.086 | 0.094 |
| Other GQ | 0.082 | 0.088 | 0.089 | 0.097 |
| Housed | 0.299 | 0.316 | 0.218 | 0.232 |
| Status Unknown (not in Census) | 0.071 | 0.136 | 0.142 | 0.203 |
| Share of HMIS users | 0.175 | | 0.825 | |
| Weighted Total | 1226 | | 5770 | |

**Sources:** LA (2004–2014) HMIS administrative data, Houston (2004–2015) HMIS administrative data, 2010 Census.
**Notes:** Table displays the weighted share of individuals who were present in an emergency or transitional shelter in HMIS data on March 30, 2010, according to HMIS records, who appeared in the 2010 Census in various GQ types or as housed. For L.A., sample consists of HMIS shelter users under Refinement 2. Where exit dates were missing in HMIS data, we imputed an exit date based on the median stay length for users of that shelter type. Bounds are calculated per methods described in the text. For L.A., the analysis is based on HMIS shelter users under Refinement 2. All shares and counts are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorised disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-008.

**Table A7**
Coverage of Census Sheltered and Unsheltered Homeless in HMIS in Los Angeles and Houston.

| Panel A: Los Angeles | | | | |
|---|---|---|---|---|
| | Sheltered | | Unsheltered | |
| | *Lower* | *Upper* | *Lower* | *Upper* |
| In HMIS Shelter during SBE | 0.361 | 0.393 | 0.085 | 0.095 |
| Excluding 3/31 exits and WSP | 0.331 | 0.359 | 0.042 | 0.046 |
| Ever in HMIS Shelter (2004–2014) | 0.681 | 0.743 | 0.336 | 0.376 |
| Weighted Total | 7344 | | 10,900 | |
| Panel B: Houston | | | | |
| | Sheltered | | Unsheltered | |
| | *Lower* | *Upper* | *Lower* | *Upper* |
| In HMIS Shelter during SBE | 0.207 | 0.218 | 0.021 | 0.022 |
| Ever in HMIS Shelter (2004–2015) | 0.720 | 0.765 | 0.623 | 0.663 |
| Weighted Total | 2515 | | 2578 | |

**Sources:** LA (CA-600, 2004–2014) HMIS administrative data, Houston (TX-700, 2004–2015) HMIS administrative data, 2010 Census.
**Notes:** Table displays the weighted share of individuals who were enumerated as sheltered and unsheltered homeless in the Los Angeles CoC who were present in HMIS shelters on March 30, 2010 ("in HMIS shelter during SBE") or ever in an HMIS shelter during the 2004–2014 period ("ever in HMIS shelter"), according to HMIS records. Where exit dates were missing in HMIS data, we imputed an exit date based on the median stay length for users of that shelter type. Lower bound assumes that the probability of being PIKed in HMIS data conditional on being PIKed in the Census is equal to one. Upper bound assumes that probability of being PIKed in HMIS data is independent of probability of being PIKed in Census data.

## References


Abowd, John M., Bell, William R., David Brown, J., Hawes, Michael B., Heggeness, Misty L., Keller, Andrew D., Mule Jr., Vincent T., et al., 2020. Determination of the 2020 U.S. Citizen Voting Age Population (CVAP) using administrative records and statistical methodology technical report. CES-20-33. CES Working Paper Series. U.S. Census Bureau, Washington, D.C.

Agans, Robert P., Jefferson, Malcolm T., Bowling, James M., Zeng, Donglin, Yang, Jenny, Silverbush, Mark, 2014. Enumerating the hidden homeless: strategies to estimate the homeless gone missing from a point-in-time count. J. Off. Stat. 30 (2), 215–229. doi:10.2478/jos-2014-0014.

Browning, M., Chiappori, P., Weiss, Y., 2014. Economics of the Family (Cambridge Surveys of Economic Literature). Cambridge University Press, Cambridge doi:10.1017/CBO9781139015882.

Corinth, Kevin C., 2015. Street homelessness: a disappearing act? AEI Paper & Studies 11 (7), 1–16.

Cronley, Courtney, 2011. A cross-level analysis of the relationship between organizational culture and technology use among homeless-services providers. Cityscape: J. Policy Dev. Res. 13 (1), 25.

Glasser, Irene, Hirsch, Eric, Chan, Anna, 2013. Ethnographic study of the group quarters population in the 2010 census: homeless populations. Center for Survey Measurement Study Series (Survey Methodology #2013-14). U.S. Census Bureau.

Hopper, Kim, Shinn, Marybeth, Laska, Eugene, Meisner, Morris, Wanderling, Joseph, 2008. Estimating numbers of unsheltered homeless people through plant-capture and postcount survey methods. Am. J. Public Health 98 (8), 1438–1442. doi:10.2105/AJPH.2005.083600.

Layne, Mary, Wagner, Deborah, 2014. The Person Identification Validation System (PVS): applying the center for administrative records research and applications' (CARRA) record linkage software. CARRA Working Paper Series #2014-01. U.S. Census Bureau, Washington, D.C.

Layne, Mary, Wagner, Deborah, Rothhaas, Cynthia, 2014. Estimating record linkage false match rate for the person identification validation system. CARRA Work. Pap. Ser., Work. Pap. #2014-02.

Martin, Elizabeth, 1992. Assessment of S-night street enumeration in the 1990 Census. Eval Rev 16 (4), 418–438. doi:10.1177/0193841X9201600407.

Meyer, Bruce D., Wyse, Angela, Corinth, Kevin, 2022. The size and census coverage of the U.S. homeless population. *NBER Work. Pap. Ser.* No. 30163.

National Research Council, 2012. Small Populations, Large Effects: Improving the Measurement of the Group Quarters Population in the American Community Survey. National Academies Press, Washington, D.C. doi:10.17226/13387.

O'Flaherty, B., 2019. Homelessness research: a guide for economists (and friends). J. Hous. Econ. 44 (C), 1–25.

Russell, Deborah, Barrett, Diane, 2013. 2010 Census Service-Based Enumeration Assessment Report. U.S. Census Bureau.

Shlay, Anne B., Rossi, Peter H., 1992. Social science research and contemporary studies of homelessness. Annu. Rev. Sociol. 18, 129–160.

Smith, Amy Symens, Charles Holmberg, and Marcella Jones-Puthoff. 2012. "The emergency and transitional shelter population: 2010," September, 22.

Smith, Annetta C., Smith, Denise, 2001. Emergency and Transitional Shelter Population: 2000. U.S. Census Bureau October.

U.S. Department of Housing and Urban Development. 2010. "The 2009 annual homeless assessment report."

U.S. Department of Housing and Urban Development. 2012. "A guide to counting sheltered homeless people (Third revision)," January.

U.S. Department of Housing and Urban Development. 2014. "Point-in-time count methodology guide."

U.S. General Accounting Office. 1985. "Homelessness: a complex problem and the federal response."

U.S. Department of Housing and Urban Development, 2018. The 2018 Annual Homeless Assessment Report (AHAR) to Congress, Part 1: Point-in-time estimates of homelessness.

U.S. Department of Education, 2021. Determining eligibility for McKinnkey-Vento rights and services. National Center for Homeless Education, Best Practices in Homeless Education Brief Series.

Wolter, Kirk M., 1986. Some coverage error models for census data. J. Am. Stat. Assoc. 81 (394), 338–346.

Wright, James D., Devine, Joel A., 1992. Counting the homeless: the census bureau's 'S-Night' in Five U.S. cities. Eval. Rev. 16 (4), 355–364. doi:10.1177/0193841X9201600401.

Zhang, Li-Chun., 2019. A note on dual system population size estimator. J. Off. Stat. 35 (1), 279–283.