NBER WORKING PAPER SERIES

THE SIZE AND CENSUS COVERAGE OF THE U.S. HOMELESS POPULATION

Bruce D. Meyer
Angela Wyse
Kevin Corinth

Working Paper 30163
http://www.nber.org/papers/w30163

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2022, Revised January 2023

This paper, which has been subject to a limited Census Bureau review, is released to inform interested parties of research and to encourage discussion. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has reviewed this data product for unauthorized 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Size and Census Coverage of the U.S. Homeless Population
Bruce D. Meyer, Angela Wyse, and Kevin Corinth
NBER Working Paper No. 30163
June 2022, Revised January 2023
JEL No. J0,R0

ABSTRACT

Despite widespread concern about homelessness, fundamental questions about the size and characteristics of this hard to study population are unresolved, in large part because it is unclear whether existing data are sufficiently complete and reliable. We examine these questions as well as the coverage of new microdata sources that are designed to be nationally representative and will allow pathbreaking new analyses. We compare three restricted use data sources that have been largely unused to study homelessness to less detailed public data. In doing this triangulation of sources, we examine the completeness and accuracy of available data and improve our understanding of the size of the homeless population and its inclusion in the Census and household surveys. Specifically, we compare restricted data from the 2010 Census, American Community Survey (ACS), and Homeless Management Information System (HMIS) to HUD's public-use point-in-time (PIT) estimates and the Housing Inventory Count (HIC) at the national, city and county, and person level. We also develop a new approach to estimating the size of the sheltered homeless population using linked Census and HMIS shelter microdata. Our analyses suggest that on a given night there are about 600,000 people experiencing homelessness in the U.S., about one-third of whom are sleeping on the streets and two-thirds in shelters. More than 90 percent of those in shelters were counted in the Census, although many were classified as housed or in other group quarters, a result that stems largely from ambiguity in the definition of a homeless shelter. By establishing the broad coverage and reliability of the new data sources, this paper lays the foundation for pathbreaking future work on the characteristics, income, safety net participation, mortality, migration, geographic distribution, and housing status transitions of the U.S. homeless population.

Bruce D. Meyer
Harris School of Public Policy
University of Chicago
1307 E 60th Street
Chicago, IL 60637
and NBER
bdmeyer@uchicago.edu

Angela Wyse
University of Chicago
1307 E 60th Street
Chicago, IL 60637
awyse@uchicago.edu

Kevin Corinth
Harris School of Public Policy
University of Chicago
1307 E 60th St
Chicago, IL 60637
kcorinth@uchicago.edu

1. Introduction

Despite widespread concern about 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 documentation and reports, and the empirical extent of coverage has not been examined. In this paper, we compare three restricted use data sources that have been largely unused to study homelessness to less detailed public data at the national, local, and person level. In doing this triangulation of sources, we provide valuable information about the size of the homeless population and its inclusion in widely used household surveys. We also develop a new approach to estimating the size of the sheltered homeless population by linking together Census and administrative shelter microdata. In doing so, we evaluate the usefulness of these data sources 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 (Glasser, Hirsch, and Chan 2014; Corinth 2015).

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 (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 a lack of HUD oversight and enforcement of methodological standards. O'Flaherty (2019) observes that PIT data on the unsheltered homeless 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, has gone largely unstudied in prior work. By comparing the PIT's estimate of the U.S. homeless population size to independent estimates at the national, local, and person level, 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 linking keys so that the microdata can be linked at the person level 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. By investigating the Census's ability to include people experiencing homelessness in its decennial count and household surveys, this paper also provides valuable insight into the completeness of some of our most fundamental sources of data on the U.S. population.

We begin by comparing unsheltered and sheltered homeless estimates at the national level. We find that the Census and PIT's unsheltered estimates are quite close to one another. Moreover, despite what appear at first to be some major differences in these sources' sheltered homeless estimates, they in fact produce similar estimates once we account for some fairly 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 the sheltered homeless population by over 30 percent in each year. Adjusting for these 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.

For our second set of analyses, we construct city and county Census homeless population estimates and use regression analysis to compare these to the 2010 PIT. We explore potential explanations for sub-national differences in the magnitude of estimates. On average, a given city or county has about three-quarters as many shelters underlying its Census estimate as it has underlying its PIT estimate. Using our regression coefficients, we estimate that differences in the number of shelters explain about one-third of the difference between the 2010 Census and PIT that remains after the definitional adjustments described above. This does not necessarily indicate, however, that the Census missed a large number of shelters. Instead, our microdata comparisons later in the paper provide strong evidence that the Census did include many of these facilities but classified them as housing units or other types of group quarters.

Our third major 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. After accounting for some 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 runs 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 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 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 individual are included on the Census questionnaire of a household where they occasionally reside or where they resided at some date 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 systems 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 found by the Census as an estimate of the share of the true sheltered homeless population found by 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. 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 wholly novel population estimate, this section serves a blueprint for future researchers to employ in estimating 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 appears 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 with our dual systems approach, which does not make assumptions about the completeness of the Census or PIT and yet arrives at a similar total. Taken together, these estimates suggest that on a given night there are about 600,000 people experiencing homelessness in the U.S., about one-third of whom are sleeping on the streets and two-thirds in homeless shelters. At the same time, our results highlight the fact that there is considerable ambiguity about what types of facilities constitute a homeless shelter 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. In an ongoing project, we link individuals from these datasets to longitudinal tax records as well as administrative data on the Supplemental Nutrition Assistance Program (SNAP), Medicare, Medicaid, Disability Insurance (DI), Supplemental Security Income (SSI), veterans' benefits, and housing assistance to provide the first national estimates of formal employment, income, and program participation based on administrative data for the U.S. homeless population. In a second project, we link the Census to the ACS to document this population's migration patterns and rates of transitions between homelessness and various housed statuses, such as traditional housing, incarceration, and other group quarters locations like group homes and residential treatment centers. In an additional project, we link individuals from these datasets to administrative data from the Social Security Administration (SSA) to provide the first national analysis of homeless individuals' mortality patterns. This research agenda has the potential to transform our understanding of this severely deprived segment of the U.S. population.

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. Section 4 describes our methodology and results for the national comparisons, while Sections 5 and 6 summarize methodology and results from the city/county and microdata comparisons, respectively. Section 7 describes our dual systems 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 services 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

[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).

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 maintenance 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, Elizabeth 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" (A. C. Smith and Smith 2001).

The 2010 SBE fared better than the previous two attempts. Meyer et al. (2021) provided a preliminary analysis of the characteristics of those included in the 2010 Census homeless counting operation and demonstrated 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" (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 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, Chiappori, and Weiss 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 homelesss 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" (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. 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 unemployed or poor at a point in time.

3. Data

This section describes our five major sources of homeless population estimates: the 2007-2020 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) used in our individual-level analysis.

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

In order to maintain federal funding, HUD requires that CoCs produce sheltered homeless population estimates every year and unsheltered estimates at least every other year. 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 shelter-use records tracked through the Homeless Management Information System (HMIS) database.

In conducting the PIT, CoCs simultaneously compile an inventory of all beds available for occupancy on the night of the PIT. This inventory is published in a separate dataset called the Housing Inventory Count (HIC). For each shelter, the HIC 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 Census counted people experiencing homelessness during its Service-Based Enumeration (SBE) operation from March 29-31, 2010. Field staff visited emergency and transitional shelters, soup kitchens, regularly scheduled mobile food vans (RSMFVs), and targeted non-sheltered outdoor locations (TNSOLs, e.g. street intersections or parks where homeless individuals were known to congregate). Unlike the PIT, the Census trained enumerators to use uniform methods and apply the same standards nationwide. They collected name and date of birth when possible, enabling researchers to link individuals from the Census to other sources.

The Census developed its list of shelters and unsheltered locations for the SBE over the course of several research and validation operations. The 2000 SBE universe served as the foundation for this list and was updated based on internet research and input from local, state, and tribal government units and homeless advocacy organizations. Census field staff visited sites during the Group Quarters Validation (GQV) and Group Quarters Advance Visit (GQAV) operations to verify their classification and prepare site administrators for the SBE.

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 non-SBE locations.

A team of outside 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, Hirsch, and Chan 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 (A. S. Smith, Holmberg, and Jones-Puthoff 2012).[2]

3.3 2006-2018 American Community Survey (ACS)

The ACS differs from the PIT and Census in that it only counts homeless people 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 through a series of operations, 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) database in Los Angeles (2004-2014) and Houston (2004-2015). Shelters that

---

[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.

receive federal funding are required to track homeless individuals' shelter use in a CoC-administered HMIS database. Some shelters that do not receive federal funding elect to use HMIS.

Shelter administrators must collect a number of universal data elements from clients, including name and date of birth, social security number, various characteristics (e.g. race, ethnicity, gender, veteran status, disabling conditions). They must also track the start and end dates of shelter enrollment. CoCs also use HMIS to track the usage of some non-shelter programs, including permanent supportive housing, rapid re-housing programs, and various other services. 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 because they in theory indicate the number of people in a subset of shelters at a point in time, although HUD instructs CoCs to ensure that entry and exit date tracking is reasonably complete and accurate before relying on HMIS-based population totals (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.

Figures 1 and 2 present estimates of the unsheltered and sheltered homeless populations for each year a given source is available.[3] In Figure 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 210,000. As seen in Figure 2, 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 about 209,000, or 52 percent of the PIT. 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

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

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 and the Census/ACS

As a first step towards reconciling differences between the PIT, Census, and ACS, we account for a handful of straightforward definitional differences. 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 (SBE) and outside the scope of the ACS's sheltered homeless estimate:

- *Domestic violence shelters.* The PIT includes people staying in shelters intended for victims of domestic violence. For the sake of privacy, such shelters are not included in the Census's Service-Based Enumeration. They are grouped instead with the Group Quarters (GQ) code for religious group quarters and are not separately identified even in restricted-use data.
- *Safe Havens.* Safe Havens are small-scale facilities offering supportive housing to hard-to-reach individuals with a history of chronic homelessness and mental illness for an unspecified length of time. As a form of supportive housing, these facilities are distinct from emergency and transitional shelters and fall outside the SBE's scope.
- *Voucher-funded hotel and motel rooms.* The PIT includes people occupying hotel/motel beds made available by a homeless service provider through vouchers or other forms of payment. 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).
- *Beds in non-shelter facilities.* The PIT also includes people occupying "beds located in a church or other facility not dedicated for use by persons who are homeless" in its

sheltered homeless count. These sites would typically not be included in 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.

We adjust the Census and ACS estimates to better align their definition of sheltered homelessness with that of the PIT at the CoC level. In the case of safe havens, it is straightforward to obtain population estimates because HUD publishes these totals at the CoC level. 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 Eliminating 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 benchmarks produced by the Census Bureau's Population Estimates Program (PEP).

For the sheltered homeless, the above-described 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 ONI population benchmark, however, is based on the most recent Census, during which this category includes several additional group quarters types that are outside the ACS's scope, namely unsheltered locations (soup kitchens, food vans, and TNSOLs), domestic violence shelters, and a few smaller categories. Figure 3 provides a graphical representation of the ONI category and the various GQ types. This means that the sheltered homeless population estimates are inflated to represent people in the broader ONI population who are not in the ACS's scope.

In order to adjust for this bias, we estimate the factor by which the ACS scales up the in-scope population in a given year, which we call $ACS\ Scaling\ Factor_t$. We call the homeless estimate that the ACS would obtain in year t if it did not conduct the above-described scaling process the $Weighted\ Homeless_{t,ACS}$ and call the sheltered estimate obtained after scaling the $Upweighted\ Homeless_{t,ACS}$. This second estimate is the one reported in Figure 2 of this paper. We therefore estimate the $ACS\ Scaling\ Factor_t$ as follows:

$$ACS\ Scaling\ Factor_t = \frac{ONI\ Population_{2010}^{Census\ Scope}}{ONI\ Population_{2010}^{ACS\ Scope}}$$

where $ONI\ Population_{2010}^{Census\ Scope}$ is the population from the 2010 Census of people in all GQ types in the ONI category, and $ONI\ Population_{2010}^{ACS\ Scope}$ is the subset of that population that is residing in GQ types that are in the ACS's scope.[4,5]

We can then use $ACS\ Scaling\ Factor_t$ to estimate the following target:

$$Weighted\ Homeless_{t,ACS} = Upweighted\ Homeless_{t,ACS} * \frac{1}{ACS\ Scaling\ Factor_t}$$

4.3 Results from the comparison of aggregate estimates

Figure 4 presents sheltered homeless population estimates with definitional and weighting adjustments. We see that the adjusted 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 estimates in 2010. Table 1 displays the year-by-year population estimates for each category of

[4] The Census does not separately identify domestic violence shelter residents from religious GQ residents, even in restricted-use data, due to privacy concerns. We must therefore estimate the share of people in the religious GQ category who are in domestic violence shelters so that we can subtract this population from the denominator. We do so under the assumption that the ratio of domestic violence to non-domestic violence sheltered homeless individuals in the Census is equal to this ratio in the PIT.

[5] In theory, the population benchmark applied to the ONI group can vary across years. However, in practice, updates to the ONI population benchmark are limited between Censuses. A 2012 report on GQ weighting methodology produced by the National Research Council observed, "As the decade progresses, the census counts become increasingly outdated and the updates … cannot always be relied on, which affects the overall quality of the GQ population estimates. For some GQ types, the population estimates are basically the decennial census counts kept constant" (National Research Council 2012).

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 ACS estimates rise by a much smaller amount because the definitional adjustment, which increases the population estimate, is almost entirely 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 1.321. In other words, we estimate that the standard ACS person weights over-weighted the homeless population estimate by about 32 percent to represent people residing in domestic violence shelters, at unsheltered locations, and in other GQ 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 a given 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 will 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.

4.4 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 3 reports the share under 18, gender/sex, race, and Hispanic ethnicity of sheltered individuals in the 2016 ACS and PIT.[6] 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). 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

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

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 4 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 our microdata comparison.

5. Comparisons of CoC-level estimates

We consider two sources of bias in the Census estimate relative to the PIT, i.e. factors which may systematically cause a CoC's Census estimate to be greater or less than the PIT estimate, specifically:[7]

- *Shelter list completeness.* We expect a more extensive shelter list in a given source to be associated with a larger population estimate. We note that the Census could have fewer shelters than the PIT because it missed shelters or because it classified them as other types of facilities and counted them outside the SBE.[8]
- *Weather and climate*. All else equal, we expect warmer temperatures and less precipitation to be associated with larger unsheltered estimates and smaller sheltered estimates in a given data source. We consider temperature and precipitation on the day of

[7] In addition, though unrelated to aggregate differences, we would also like to know whether certain observed aspects of PIT methodology produce more reliable estimates on average. We therefore consider various factors that could explain imprecision in CoC-level estimates, defined as the absolute percentage difference between the Census and PIT estimates. We present the results from that exercise in Section A1 of the appendix. This analysis may prove useful for other researchers conducting CoC-level analyses using the PIT.

[8] We discuss the possibility of ambiguous facility classification at length in Section V.

the Census and PIT as a measure of weather, and average temperature and precipitation in the month of the Census and PIT as a measure of climate.

As a robustness check, we also consider possible relationships between relative bias and CoC characteristics, including the share of the CoC's population in urban areas, the share in poverty, the Median Rent Index (MRI), and the CoC's population density. It is not clear, however, that we should expect a particular pattern of relationships between these variables and relative bias.

5.1 Modeling the relative size of estimates[9]

We define relative size as the ratio of the Census to the PIT estimate and use OLS to estimate the following models:[10]

$$\frac{S_i^{Census}}{S_i^{PIT}} = \beta_0 + \beta_1 \frac{Shelters_i^{Census}}{Shelters_i^{PIT}} + Weather_i'\beta_3 + Characteristics_i'\beta_4 + \epsilon_{1,i}$$

$$\frac{U_i^{Census}}{U_i^{PIT}} = \gamma_0 + Weather_i'\gamma_1 + Characteristics_i'\gamma_2 + \epsilon_{2,i}$$

where S_i (U_i) is CoC i's sheltered (unsheltered) homeless count, $Shelters_i$ is the number of unique shelter addresses in the CoC, $Weather_i$ is a vector of temperature and precipitation variables indicating temperature and precipitation on the day of and in the month of each count, and $Characteristics_i$ is a vector of CoC characteristics.[11]

We estimate the number of shelters in the Census as the number of unique addresses associated with sheltered homeless individuals in the boundaries of a given CoC. We estimate the number of shelters in the 2010 PIT using the number of unique housing inventory records in

[9] For this section's analyses, we first construct CoC-level sheltered and unsheltered population estimates in the Census. We use HUD files indicating the coordinates of CoC boundaries to match Census block groups to CoCs, which in turn allows us to assign homeless individuals in the Census to CoCs based on the block group associated with their address identifier.
[10] We use Weighted Least Squares (WLS) in our estimations to account for the likely presence of heteroskedasticity in our variables relative to CoC population size.
[11] We consider three different models for weather and climate variables: one that includes the levels of temperature and precipitation, one that adds squared terms for weather and climate variables, and one where the weather and climate coefficients are constrained to be equal, but opposite in sign in the Census and the PIT.

the HIC listed for facilities classified as shelters.[12] We draw weather information from the National Climatic Data Center (NCDC)'s Global Historical Climatology Network (GHCN) daily summaries datasets and average daily and monthly weather readings across all weather stations located within a given CoC (Menne et al. 2012).

Because we wish to explain aggregate differences between these sources, we exclude CoCs that are so small as to not make a meaningful contribution to aggregate differences. For the sheltered model, we limit our sample to CoCs where the minimum of the PIT or Census is greater than or equal to 100. For the unsheltered, we limit our sample to those where the minimum is greater than or equal to 50.

5.2 Results from comparison of CoC-level estimates

We find a positive, statistically significant association between the Census-PIT ratio of sheltered estimates and the Census-PIT ratio of unique shelter addresses, meaning that CoCs with more shelters in the PIT or Census tend to have higher sheltered homeless estimates in that source.[13] Specifically, a one percent increase in the number of shelters in the Census is associated with a 0.395 percent increase in the Census sheltered homeless estimate, holding the PIT shelter count constant. Equalizing the number of unique shelter addresses in each source – i.e. setting the mean shelter count ratio equal to one – would be predicted to increase the Census-to-PIT sheltered population estimate ratio from 0.77 to 0.85, closing approximately one-third of the relative bias In sheltered estimates that remains after aligning the definition of homelessness between sources.[14]

We also observe that weather and climate variables (temperature and precipitation on the day of and month of the Census and PIT counts) are jointly insignificant at the five percent level.[15] CoC population characteristics (share urban, share poor, population density, and Median

[12] We are concerned about two potential sources of bias in this measure of the number of PIT shelter addresses. While the HIC does not contain addresses in 2010, we observe in later years that multiple housing inventory records can be associated with a single address. Similarly, some inventory records are labeled as being "multi-site," meaning that they can in theory be associated with multiple addresses. These sources of bias work in opposite directions. We apply a correction to the number of PIT shelters based on estimates of the magnitude of each source of bias in later years. As a robustness check, we also present our findings without this bias correction.
[13] Table A1 in the Appendix displays key results from this estimation.
[14] Table A9 displays the relevant summary statistics for these regressions.
[15] Table A7 in the Appendix displays results from the hypothesis tests described in this section.

Rent Index) are also jointly insignificant at the five percent level. Seasonality and CoC characteristics do not appear to explain relative bias in CoC-level sheltered estimates.

Seasonality appears to explain some of the variation in the relative magnitude of unsheltered estimates, but the observed signs do not always accord with the most intuitive predictions.[16] Variables indicating temperature and precipitation on the day of a count and averaged across the month of the count are jointly significant. Some of the signs on these variables are consistent with our hypotheses, while others are not. One variable – average daily precipitation in the month of the PIT – is statistically significant and negative, even though we expect this variable to be associated with a larger unsheltered Census estimate relative to the PIT. We therefore note that seasonality seems to matter for the relative magnitude of the unsheltered count estimates but are hesitant to place too much emphasis on any particular coefficient estimates given the puzzling pattern of results.[17] CoC characteristics do not appear to explain relative bias in CoC-level unsheltered estimates; these variables are jointly insignificant at the five percent level.

Taken together, these results suggest that there is a fair amount of error in both the Census and PIT estimates and that these counts are of uneven quality across geographies, particularly in the case of unsheltered estimates. This pattern suggests that there would likely be little bias in regressions with the PIT count as the dependent variable but using this count as an explanatory variable would lead to substantial bias.

6. 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 the Los Angeles and Houston CoCs to the 2010 Census using Protected Identification Keys (PIKs) available on both sources. These links allow us to observe whether and in what housing

[16] Table A2 in the appendix displays key results from our unsheltered relative bias estimation.
[17] Table A5 indicates the effect on aggregate Census unsheltered estimates and the mean Census/PIT ratio that we would expect if weather and climate on the day of the Census were equal to that of the PIT. Those calculations suggest that the aggregate unsheltered Census would change by -8.0 to 5.5 percent and that the mean Census/PIT ratio would change by -34.0 to -6.3 percent. The predicted effect is very sensitive to the model chosen (i.e. level or level and squared terms) and the outcome considered, and thus do not paint a clear picture of seasonality's overall effect on unsheltered estimates.

status particular individuals from HMIS data were included in the Census. Because HMIS data are a key input to the PIT, this approach proves informative about the coverage and accuracy of both the Census and PIT.

6.1 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 SBE. In this sub-section, we assess the accuracy of shelter entry and exit dates in HMIS and describe our approach to overcoming the problems we document.

Evaluating the accuracy of HMIS shelter entry and exit dates

Figure 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 Housing Inventory Count (HIC). In Los Angeles, capacity increases each winter as part of the city's Winter Shelter Program, which runs from December 1 to March 15. Because the HIC indicates beds available on a single date each year, we extrapolate linearly from one year's total 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. Daily occupancy, as implied by recorded entry and exit dates, nearly doubles between December and March of 2009-2011, compared to much more moderate seasonal growth in the winters of 2012 and 2013. We also observe precipitous drops on a handful of days, including March 31 of 2009 and 2010 and June 15 of 2011, although we are not aware of any reason why so many people would exit shelters on these particular days.[18] We suspect that HMIS administrators conducted a "purge" of open shelter spells on those dates. Analyses of shelter entry rates (per one thousand L.A. residents) and hazard rates for shelter exit (i.e., probability of exit in a given month conditional on being in

[18] 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.

the shelter at the beginning of that month) suggest that the above-described patterns are driven by incorrect exit dates, not incorrect entry dates.[19]

Figure 6 displays daily occupancy and capacity in Houston HMIS data for 2009-2013. Unlike in Los Angeles data, we do not observe precipitous 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 obvious and systematic fashion than in Los Angeles HMIS data.

Additional evidence on errors in HMIS entry or exit date reporting

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 in HMIS, implying an erroneous entry or exit date for at least one of the spells. Moreover, using methods that we describe shortly, 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.[20,21]

Refinements to Los Angeles HMIS data

In subsequent sections, we use HMIS data to identify people who were in shelters during the Service-Based Enumeration (SBE). Given our assessment of exit date quality in Los Angeles, we make a series of refinements to HMIS data to drop individuals who we suspect were

[19] Table A11 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 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.
[20] 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.
[21] 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.

erroneously indicated by HMIS as being in a shelter on that date. The first refinement drops people with an exit date of March 31, 2010, as Figure 5 suggested a purge of open spells on that date. The second refinement further drops people recorded in a shelter with a shelter name indicating participation in the Winter Shelter Program, which ended on March 15, 2010, thereby giving us confidence that these shelter spells concluded prior to the SBE at the end of March. Finally, the third refinement drops people with entry dates prior to March 1, in accordance with our assessment that entry dates are more reliable than exits, combined with the fact that most spells are short.

6.2 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 Identification Numbers (ITINs) and other information (Wagner and Layne 2014).

Table 5 presents the share of records in our HMIS and Census datasets that are assigned a PIK by the Census. 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 PIK. Census data do not contain SSNs, so linkage rates depend on the completeness and accuracy of personally identifiable information provided to enumerators, the uniqueness of this information, and the coverage of the reference file. Linkage rates for the Census data vary by SBE 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 at soup kitchens, and 17.2 percent for individuals at TNSOLs. In the next section, we describe the steps we take to adjust our estimates for non-linkage.

6.3 Methods for estimating the share of HMIS shelter users recorded in the 2010 Census

We wish to estimate the share of people indicated by HMIS data as being in emergency or transitional shelters during the SBE who appear in the 2010 Census as sheltered homeless and in other statuses. More formally, we estimate the following target:

$$HMIS\ Shelter\ User\ Share\ in\ Status\ X\ in\ Census = \frac{HMIS\ Shelter\ Users\ in\ Status\ X\ in\ Census}{All\ HMIS\ Shelter\ Users}$$

where "*All HMIS Shelter Users*" is the number of individuals who were present in an HMIS shelter according to the CoC's records on March 30, 2010, and Status *X* represents those individuals' enumeration site in the Census (e.g. homeless shelter, unsheltered location, other group quarters type, housed).[22,23] We estimate the share with unknown status (i.e. not in the Census) by subtracting the sum of the above shares from one.

We cannot directly calculate the numerator of our target, however, because we can only count the number of individuals who appeared in a given Census status and were assigned a PIK in both sources. We miss individuals who actually appeared in both sources but were not assigned a PIK in one source or both.

We calculate the numerator of our target using inverse probability weights (IPWs):

$$HMIS\ Shelter\ User\ Share\ in\ Status\ X\ in\ Census = \frac{\sum_{i \in P} I\{Status\ X\ in\ Census_i\}\omega_i}{All\ HMIS\ Shelter\ Users}$$

where set P includes all HMIS shelter users who were assigned a PIK in both the Census and HMIS data, $I\{Status\ X\ in\ Census_i\}$ is an indicator for whether individual *i* who was in an

[22] The Census's Service-Based Enumeration (SBE) took place over a three-day window, from March 29-31. Shelter administrators were able to indicate their date of preference to Census staff. We take as our SBE reference day the midpoint of the SBE, March 30. Because we think a very small share of HMIS shelter users exit or enter on a given day, we expect that this choice will introduce a minimal amount of error into our estimates.

[23] In some cases the Census records an individual as being in multiple statuses. The most frequent combination, by far, is for someone to be recorded as both housed and in a homeless shelter. We believe this arises because the Census did not de-duplicate individuals between shelters and the universe of housing units. We resolve any duplicate records in the following order of preference: sheltered homeless, unsheltered homeless, other GQ, and finally housed. Table A5 in the Appendix indicates the share of HMIS shelter users with multiple statuses in Census data

HMIS shelter on March 30, 2010 was in status X in the Census, and ω_i is the inverse of the probability that individual *i* is PIKed in both the Census and HMIS data.

This inverse probability weighting procedure requires us to estimate the probability that an HMIS shelter user is assigned a PIK in both the Census and HMIS data conditional on observed characteristics. Our data, however, 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 PIK. We would need to distinguish these two events in order to model the joint probability of being PIKed in both sources.

To address the above challenge, we devise a method to estimate bounds on ω_i. Our bounds rely on the assumption that being PIKed in one source does not make an individual less likely to be PIKed in the other source, i.e.:

$$Pr_i(PIKed\ in\ HMIS \mid PIKed\ in\ Census) \geq Pr_i(PIKed\ in\ HMIS)$$
$$Pr_i(PIKed\ in\ Census \mid PIKed\ in\ HMIS) \geq Pr_i(PIKed\ in\ Census)$$

Given the above assumptions, we can obtain a lower bound on the joint probability of being assigned a PIK in both sources in the following manner:

$$\begin{aligned} Pr_i(PIKed\ in\ HMIS\ \ and\ PIKed\ in\ Census) &= Pr_i(PIKed\ in\ HMIS \mid PIKed\ in\ Census)\, Pr_i(PIKed\ in\ Census) \\ &\geq Pr_i(PIKed\ in\ HMIS)\, Pr_i(PIKed\ in\ Census) \end{aligned}$$

Thus, the product of the individual probabilities provides a lower bound on the joint probability of being assigned a PIK in both sources.[24]

We can similarly calculate an upper bound on the joint probability of being assigned a PIK in both sources as follows:

$$\begin{aligned} Pr_i(PIKed\ in\ HMIS\ and\ PIKed\ in\ Census) &= Pr_i(\ PIKed\ in\ Census \mid PIKed\ in\ HMIS)\, Pr_i(PIKed\ in\ HMIS) \\ &\leq\ \ Pr_i(PIKed\ in\ HMIS) \\ Pr_i(PIKed\ in\ HMIS\ and\ PIKed\ in\ Census) &= Pr_i(\ PIKed\ in\ HMIS \mid PIKed\ in\ Census)\, Pr_i(PIKed\ in\ Census) \\ &\leq\ \ Pr_i(PIKed\ in\ Census) \end{aligned}$$

[24] Alternatively, we can write:

$$\begin{aligned} Pr_i(PIKed\ in\ HMIS\ \ and\ PIKed\ in\ Census) &= Pr_i(PIKed\ in\ Census \mid PIKed\ in\ HMIS)\, Pr_i(PIKed\ in\ HMIS) \\ &\geq Pr_i(PIKed\ in\ Census)\, Pr_i(PIKed\ in\ HMIS) \end{aligned}$$

Either of these individual probabilities can serve as an upper bound on $Pr_i(PIKed\ in\ HMIS\ and\ PIKed\ in\ Census)$ and whichever is smaller will provide a tighter bound. For each individual we take the probability that is smallest and use this to calculate our upper bound on $Pr_i(PIKed\ in\ HMIS\ and\ PIKed\ in\ Census)$.[25] We can then calculate upper and lower bounds on ω_i by taking the inverse of the lower and upper bounds, respectively, of $Pr_i(PIKed\ in\ HMIS\ and\ PIKed\ in\ Census)$.

We estimate these bounds using a probit model of PIK status on observed characteristics in each data source. The 2010 Census model includes indicator variables for various age and race categories, Hispanic ethnicity, sex, state dummy variables, and an indicator for residing in an urban area. For all demographic characteristics, we include interactions with indicators for whether the variable was imputed. For homeless individuals, we further include variables indicating the enumeration site (e.g. shelter, soup kitchen, etc.). For housed individuals, we also include variables indicating the building structure type (e.g. one- or multi- family housing, mobile home), tenure type (e.g. owned, rented, occupied without rent), and household type (e.g. husband-wife family household, other family household, non-family household). The HMIS model includes indicator variables for various age and race categories, Hispanic ethnicity, sex, HMIS program type (e.g. emergency shelter, transitional housing, other types of services), number of HMIS enrollment spells over the data period, average spell length, and year of first enrollment.

6.4 Results on coverage of HMIS shelter users in the Census

6.4.1 Share of HMIS shelter users in Census statuses in Los Angeles

Table 6 displays estimated lower and upper bounds on the share of HMIS shelter users counted in various statuses in the Census in Los Angeles. The first two columns display the Census coverage of HMIS shelter users when we use all HMIS records for emergency and

[25] On average, individuals are more likely to be PIKed in HMIS data than in the Census, because HMIS collects social security numbers for most users. In publicly disclosed data, we know that the probability of being PIKed in HMIS data is about 90%, and the probably of sheltered homeless individuals being PIKed in the Census is about 60%. In the vast majority of cases, we therefore use the probability of being PIKed in the Census to obtain a lower bound on ω_i.

transitional shelter spells. We then present results under three sets of refinements to the HMIS data, where each refinement is 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 L.A. 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.

We note that the share recorded as sheltered in the Census increases with each refinement, suggesting that we have succeeded in better identifying people who were truly in shelters during the SBE. 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 a large number 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 derived above. 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 status in some status or another.

6.4.2 Share of HMIS shelter users in Census statuses in Houston

The last two columns of Table 6 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 as we did with the Los Angeles data. 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 GQs. 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.

6.4.3 Explanations for HMIS/Census status discrepancies

We explore several potential reasons for discrepancies in individuals' statuses between HMIS data and the Census, including: Census classification of certain HMIS shelters as housing units; Census classification of certain HMIS shelters as other GQs; discrepancies arising from the timing of Census responses; and residual HMIS exit date errors.[26]

6.4.4 Census classification of some HMIS shelters as housing units

Ambiguity in the Census definitions of shelters could explain why some HMIS shelter users were recorded by the Census in statuses other than sheltered homeless. In particular, the HMIS definition of transitional housing leaves open the possibility that Census would classify some of these facilities as housing units. HUD defines transitional housing as programs that "provide people experiencing homelessness a place to stay combined with supportive services for up to 24 months" (U.S. Department of Housing and Urban Development 2018). Transitional housing is designed to provide homeless individuals with interim stability and support to successfully move to permanent housing. HUD requires that transitional housing residents have a lease, sub-lease, or occupancy agreement.

Census, on the other hand, does not distinguish between emergency shelters and transitional housing in its shelter definition. It defines emergency and transitional shelters as "facilities where people experiencing homelessness stay overnight… [including] shelters where

[26] We have also considered the possibility that incorrect PIKing resulted in the observed discrepancies. Table A10 in the appendix displays the share of Los Angeles HMIS shelter users enumerated in various statuses in the Census who were found in 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 results suggest that incorrect PIKing may not be a major source of status discrepancies between sources.

people know they have a bed for a specified period of time" (A. S. Smith, Holmberg, and Jones-Puthoff 2012).

Table 7 offers evidence on the possible Census classification of some HMIS transitional shelters as housing units. These tables display estimates of the coverage of Los Angeles and Houston HMIS shelter users in the 2010 Census by program type, e.g., emergency shelter versus transitional housing. In Los Angeles, about 35-37 percent of people enrolled in HMIS transitional housing facilities are recorded by the Census as being housed, compared to just 17-19 percent of those in HMIS emergency shelters.[27] We observe a similar pattern in Houston, where 26-27 percent of transitional housing occupants were recorded by the Census as housed, compared to just 14-15 percent of those in emergency shelters.

Table 8 further explores potential Census classification of transitional shelters as housing. Our goal with this table is to identify the extent to which Census classified entire transitional housing facilities as housing units. The table displays the share of HMIS shelter users in Los Angeles recorded in various Census statuses, broken down into those who are in HMIS facilities where the share recorded in the Census as sheltered ranged from 0, 0-0.5, and 0.5-1, disaggregated by HMIS shelter type.

Looking at column 5, we see that about 850 of the 1,800 people in HMIS transitional housing facilities were in facilities where no PIKed individual was recorded by the Census as being in a shelter. In column 2, we further see that about 54-55 percent of people in these facilities were recorded in the Census as being housed. That share is just 15-16 percent for people in HMIS emergency shelters. In other words, transitional shelter occupants who were recorded by the Census as housed tended to reside in facilities where *no one* was recorded by the Census as being in a homeless shelter. This lends further evidence to the hypothesis that these facilities were simply classified differently by the two sources.

6.4.5 Census classification of HMIS shelters as other GQs

[27] We also observe that about 95-100 percent all of the HMIS transitional housing occupants were recorded by the Census in some status, compared to approximately 78.8 percent of HMIS emergency shelter occupants. We discuss this pattern later in the paper when we examine the characteristics of people missed by the Census.

Table 9 explores the possibility that the Census classified certain HMIS shelters as GQ types other than homeless shelters. In refinement 2 of the Los Angeles data, we see that of the 7-8 percent of HMIS shelter users recorded by Census as being in other GQs, about 43 percent were recorded in residential treatment centers for adults. According to the Census definition, these centers “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 HMIS shelter users are recorded in this status by the Census.

In the Houston column of Table 9, we see that of the about 15-16 percent of HMIS shelter users recorded in other GQ types, approximately half of them were record by the Census as being in residential treatment centers, while another quarter were recorded as being in group homes intended for adults. This latter category is defined as “community-based group living arrangements that… provide room and board and services, including behavioral, psychological, or social programs.” About 19 percent were recorded in correctional facilities intended for juveniles.

6.4.6 Discrepancies arising from the timing of Census responses

Another possible explanation for discrepancies in individuals’ HMIS and Census statuses lies in the timeline of Census responses from housing units. The SBE was conducted March 29-31, 2010, and individuals were recorded based on their status on that date. The housing unit (HU) questionnaire distributed by mail asked respondents to list people who were “living our staying in this house, apartment, or mobile home on April 1, 2010.”

A very small number of individuals may have truly transitioned from shelters to housing between March 29-31 and April 1. 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 two-thirds had mailed these questionnaires back to the Census Bureau. The window of possible response also extended well beyond March 30. Nearly one-third of all households returned their questionnaires after April 1, including about 20 percent of all households during the Non-Response Follow-Up (NRFU) operation that began on May 1.

Using the distribution of Census response dates and shelter entry/exit patterns, in conjunction with the distribution of Census response dates obtained from various Census press releases and an official report on the 2010 Census Non-Response Follow-Up operation, 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.[28] 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.

6.4.7 Residual HMIS exit date errors

Some portion of the HMIS/Census status discrepancies can likely be attributed to remaining errors in HMIS exit dates. Table 9 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 GQs 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.

6.4.10 Caveat on the Calculation of Housing Status Probabilities

As a caveat on the preceding sections, we consider the key assumptions underlying our results and the implications of plausible deviations from those assumptions. We have assumed that being assigned a PIK is random conditional on the covariates in our inverse probability weighting (IPW) model. This means that the probability of being assigned a PIK is assumed to be the same for a randomly chosen housed or other group quarters residing person as it would be for someone who was recently in an HMIS homeless shelter but is now housed or in another group quarters, given their covariates. If instead these recently homeless individuals were less

---

[28]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 a given date range as the population as a whole.

likely to be assigned a PIK, then our IPW weights would be under-weighting those individuals. This tendency could explain our observation that the share with unknown status (the residual category) decreases with each sample refinement, if each refinement disproportionately drops underweighted rather than correctly weighted individuals.

Recently homeless individuals who transition to housing could be difficult to link for various reasons. For one thing, recently homeless individuals are probably less likely to be associated with their current address in the reference files used for PIKing, which could make it more difficult for the Census Bureau to link them.[29] Recently homeless individuals may also be more reluctant or unwilling to provide personal information on the Census questionnaire, or they may be more likely to have a tenuous attachment to their living situation which could mean that the person responding to the Census questionnaire lacked complete or accurate information for them. About 90 percent of all housed people in the Census were assigned linkage keys, compared to 68 percent of the sheltered homeless. We therefore expect that 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 with unknown status.

6.5 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 10 displays weighted counts of HMIS shelter users from Los Angeles and Houston whose PIK appears more than once in various combinations of Census statuses.[30] In Los Angeles, about 800-1,000 people, or 14-17 percent of the 5,800 HMIS shelter users under Refinement 2, 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 have a duplicate record.

Table 11 examines duplication among all individuals counted in homeless statuses in the 2010 Census more broadly. Specifically, this table shows the share of all people counted in

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

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

homeless shelters and in each of the unsheltered statuses in the Census who have at least one housed record and at least one other GQ record in addition to their homeless record, as indicated by the presence of additional records with that same PIK. We estimate that about 21-24 percent of the sheltered homeless, 45-56 percent of those 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 had some other GQ record in addition to their homeless record. Among those with records in other GQ facilities, the most common facility types were group homes, residential treatment centers, state prisons, and local jails.

By way of context, it is important to note that duplication is a non-trivial issue in the Census more broadly. The 2010 Census Coverage Measurement (CCM) study found that the 2010 Census had about 8.5 million erroneous estimations due to duplication, or about 2.8 percent of all people counted that year. A report from the Department of Commerce's Office of the Inspector General (OIG) described a "high risk of duplication" for homeless individuals in particular, which they attribute to official guidance that instructed enumerators to count homeless individuals even if they stated they had been previously counted. (The report also notes that this guidance was often ignored by enumerators, who chose not to count people who stated they already had been included in the Census, U.S. Department of Commerce 2011).

To understand the source of double-counting, we first explore the possibility of erroneous linkage.[31] Table 12 displays agreement rates for age, gender, race, Hispanic ethnicity, and county and state of residence among duplicate record pairs in the Census. Among records pairs where a given characteristic is non-imputed for both records, the individuals' sex matches in about 94 percent of cases. The individuals' age matches exactly in about 78 percent of cases and within five years in about 90 percent of cases. Race and Hispanic ethnicity matched in 85 and 89 percent of cases, respectively, while in 89 percent and 81 percent of cases the two records were located in the same state and county. We consider in particular the cases where sex does not match across records as possible instances of erroneous linkage. Indeed, we see in Table 12 that among different-sex duplicates, age, race, and ethnicity agreement rates are much lower than

[31] 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, Wagner, and Rothhaas (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.

among same-sex duplicates. However, we note that these cases correspond to only about 3,500 individuals. Thus we do not interpret these results as broadly suggestive of erroneous linkage, although it is possible that errors do occasionally occur.

Beyond agreement rates for observed characteristics, several other facts give us confidence that duplication does not reflect widespread erroneous linkages. For one thing, we observe high rates of duplication even for HMIS shelter users who were counted in the Census. We have high confidence in the quality of links assigned to HMIS shelter users because those 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 (Meyer et al. 2021). We would not expect to see these patterns if the PIKs of housed individuals were erroneously assigned to homeless individuals.

Misclassification offers another potential explanation for duplicate records. It is possible that Census enumerators classified individuals observed in soup kitchens and food vans, in particular, as homeless when in fact those individuals were housed but happened to be using homeless services. However, we think 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 specifically on the basis of where they spent the night. Yet the fact that we still see a large degree of duplication for these categories suggests that misclassification does not explain the majority of duplicate records.

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, frequent transitions between homelessness and housing during the long window of Census response are likely to occur and could lead to double-counting. Moreover, the 2010 Census questionnaire instructed respondents to count all people “who live and sleep here most of the time.” It is therefore possible that some homeless individuals might be counted at the residence of a relative or acquaintance where they sometimes reside, despite not having been there on the Census reference date of April 1 specifically.

We explore this possibility in Tables 13. Table 13 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 (either adult or minor) 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 that leads us to strongly suspect that double-counting arises primarily from these individuals' inclusion on a family member or acquaintance's Census form.

7. Dual Systems 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 the under coverage of Decennial Censuses since 1980.[32] 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 will be consistent if the number of individuals found in both samples is large and certain assumptions are met.

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

As an equation

$$\text{(1)}\quad \mathit{Sheltered\ Homeless\ Estimate} = \mathit{Census\ sheltered\ homeless\ count} * \frac{\mathit{HMIS\ sheltered\ homeless\ count}}{\mathit{Count\ of\ HMIS\ sheltered\ homeless\ also\ found\ by\ Census\ in\ HMIS\ shelter}}.$$

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 equation (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, e.g. some transitional shelters and group homes. We provided evidence that the two sources classify some facilities differently in Section 6. We will be using equation (1) to account for this difference. Such individuals are appropriately included in the numerator of the ratio but not in the denominator. Equation (1) will also account 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 6 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 6, 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 identify the HMIS observations (or at least their count) that are from the time period outside that of the Census homeless counting operation.

To do this, we estimate the share of those recorded erroneously in HMIS that is consistent with the count found in jail or prison. We obtain this estimate 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 using the Census statuses of the sample of those who we identified as having date errors in Section 6, a group that we call HMIS shelter exiters. As an equation

(2)

$$\frac{\text{Count recorded erroneously in HMIS shelter}}{\text{HMIS sheltered homeless count}} = \frac{\text{Count found in jail or prison}}{\text{HMIS sheltered homeless count}} * \frac{\text{Count of HMIS shelter exiters}}{\text{Count of HMIS shelter exiters in jail or prison}}.$$

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[33] and multiply this by the estimated count recorded erroneously in an HMIS shelter obtained using equation (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 14 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 equation (2), we estimate that about 36-38 percent of the HMIS sheltered homeless were erroneously recorded in an HMIS shelter. 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

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

estimate of 209,000 as in equation (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.

7.1 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.[34] Applying these assumptions to our setting, the dual systems estimator from equation (1) will provide a consistent estimate of the true sheltered homeless population if the following conditions are met:

1) There are no duplicated records or erroneous enumerations in either the HMIS or the Census homeless count;
2) The matched records between the HMIS and Census counts can be identified without errors;
3) The average HMIS capture probability for people in our Census dataset is equal to the average HMIS capture probability for sheltered homeless individuals not in our Census dataset;
4) Captures in the HMIS are uncorrelated with one another (aside from intra-cluster correlations, which are permitted).[35]

To address the first assumption, we deduplicate records using PIKs 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 fairly confident that the first assumption is

[34] By treating the administrative list as fixed, this approach circumvents the problem of modeling the popuation 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).
[35] 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 systems estimator.

reasonably close to true. 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 is difficult to test, but 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 without violating this assumption).

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.[36] 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.[37] This remains a caveat on our findings and a potential question for future work linking other CoCs' HMIS data to the Census.

8. Discussion

8.1 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 estimates at the national, city and county, and person level and by using dual systems methods to obtain a new estimate of the sheltered homeless population that is reliable under fairly 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 PIT relative to the true homeless population. We discuss how these sources of bias could

[36] See table A11 in the Appendix.
[37] 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.

affect aggregate comparisons and how they might explain differences between the PIT and Census's sheltered homeless estimates and the dual systems estimate.

8.1.1 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 each source relative to the true population, and we are unable to estimate this population using dual systems 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 SBE or PIT) 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:

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

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. Our CoC-level comparisons accounting for several measures of temperature and precipitation suggest 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 10 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 could 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.

8.1.2 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 in 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 systems methodology described in Section 7, 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 good estimate of the sheltered homeless population size. 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 systems 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 our dual systems estimate is lower than the PIT estimate.

Double counting, on the other hand, is less of a concern in sheltered counts 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 the collection of personal information was far from complete preventing comprehensive deduplication. Personal information was collected for most of those in the Census sheltered count, reducing the probability of double counting due to non-linkage in the Census.

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 our aggregate comparisons, our micro data comparisons suggested that more subtle differences in classification were likely to be present. The combination of Census undercounting and subtle classification differences likely explains why the adjusted Census estimate falls short of the dual systems estimate. We note, however, that our dual systems estimate corrects for both of these sources of undercounting to potentially produce a reliable estimate of the true sheltered homeless population.

8.2 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 the sheltered homeless in the 2010 Census was surprisingly good. Our dual systems estimates 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 7, means that the true share could be even higher. About 61-64% were found by the Census in shelters, 19% in housing units, and 9% in other types of group quarters. The remaining 4% appear to have been misclassified as unsheltered.

As documented in Sections 6, 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, it 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 traditional housing, likely because the people residing there had occupancy agreements of up to two years. We also saw that the Census appears to have classified some HMIS facilities not as homeless shelters but as group homes for adults or residential treatment centers for substance abuse. This means 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 fact that there is no 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 the winter months and then 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.

9. 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-systems methods used in this paper may prove useful to other researchers looking to estimate the true unsheltered homeless population size, although doing so will require a set of linkable data on the unsheltered population that satisfies the assumptions described in Section 7. 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. We find that that a substantial share were in fact counted twice. 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 populations, because we find evidence of transitions between housing statuses for this set of individuals even within datasets designed to convey a static picture of the U.S. 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 have laid the foundation for pathbreaking work in progress using these data sources to learn about the demographic characteristics, income, 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.

9. Citations

Agans, Robert P., Malcolm T. Jefferson, James M. Bowling, Donglin Zeng, Jenny Yang, and Mark Silverbush. 2014. "Enumerating the Hidden Homeless: Strategies to Estimate the Homeless Gone Missing From a Point-in-Time Count." *Journal of Official Statistics* 30 (2): 215–29. https://doi.org/10.2478/jos-2014-0014.

Burt, Martha R, Cohen, Barbara E. 1989. "America's Homeless: Numbers, Characteristics, and Programs That Serve Them." *Urban Institute Press*.

Browning, Martin, Pierre-Andre Chiappori, and Yoram Weiss. 2014. "The Economics of the Family." Cambridge: Cambridge University Press. doi:10.1017/CBO9781139015882.

Conroy, Stephen J., and David M. Heer. 2003. "Hidden Hispanic Homelessness in Los Angeles: The 'Latino Paradox' Revisited." *Hispanic Journal of Behavioral Sciences* 25 (4): 530–38. https://doi.org/10.1177/0739986303258126.

Corinth, Kevin C. 2015. "Street Homelessness: A Disappearing Act?" *AEI Economic Perspectives*, June.

Cronley, Courtney. 2011. "A Cross-Level Analysis of the Relationship Between Organizational Culture and Technology Use Among Homeless-Services Providers." *Cityscape: A Journal of Policy Development and Research* 13 (1): 25.

Culhane, Dennis P., Edmund F. Dejowski, Julie Ibañez, Elizabeth Needham, and Irene Macchia. 1994. "Public Shelter Admission Rates in Philadelphia and New York City: The Implications of Turnover for Sheltered Population Counts." *Housing Policy Debate* 5 (2): 107–40. https://doi.org/10.1080/10511482.1994.9521155.

Glasser, Irene, Eric Hirsch, and Anna Chan. 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.

———. 2014. "Reaching and Enumerating Homeless Populations." In *Hard-to-Survey Populations*, edited by Roger Tourangeau, Brad Edwards, Timothy P. Johnson, Kirk M. Wolter, and Nancy Bates, 180–200. New York, NY: Cambridge University Press.

Hopper, Kim, Marybeth Shinn, Eugene Laska, Morris Meisner, and Joseph Wanderling. 2008. "Estimating Numbers of Unsheltered Homeless People Through Plant-Capture and Postcount Survey Methods." *American Journal of Public Health* 98 (8): 1438–42. https://doi.org/10.2105/AJPH.2005.083600.

James, Franklin J. 1991. "Counting Homeless Persons with Surveys of Users of Services for the Homeless." *Housing Policy Debate* 2 (3): 733–53. https://doi.org/10.1080/10511482.1991.9521071.

Martin, Elizabeth. 1992. "Assessment of S-Night Street Enumeration in the 1990 Census." *Evaluation Review* 16 (4): 418–38. https://doi.org/10.1177/0193841X9201600407.

McCallum, Hamish. 2000. "Population Parameters: Estimation for Ecological Models." Oxford; Malden, Mass.: Blackwell Science.

Menne, Matthew J., Imke Durre, Russell S. Vose, Byron E. Gleason, and Tamara G. Houston. 2012. “An Overview of the Global Historical Climatology Network-Daily Database.” *Journal of Atmospheric and Oceanic Technology* 29 (7): 897–910. https://doi.org/10.1175/JTECH-D-11-00103.1.

Metraux, Stephen, Dennis Culhane, Stacy Raphael, Matthew White, Carol Pearson, Eric Hirsch, Patricia Ferrell, Steve Rice, Barbara Ritter, and J Stephen Cleghorn. 2001. “Assessing Homeless Population Size Through the Use of Emergency and Transitional Shelter Services in 1998: Results from the Analysis of Administrative Data from Nine US Jurisdictions.” *Public Health Reports* 116: 9.

Metraux, Stephen, Nicholas Eng, Jay Bainbridge, and Dennis P. Culhane. 2011. “The Impact of Shelter Use and Housing Placement on Mortality Hazard for Unaccompanied Adults and Adults in Family Households Entering New York City Shelters: 1990–2002.” *Journal of Urban Health* 88 (6): 1091–1104. https://doi.org/10.1007/s11524-011-9602-5.

Metraux, Stephen, Janna Manjelievskaia, Dan Treglia, Roy Hoffman, Dennis P. Culhane, and Bon S. Ku. 2016. “Posthumously Assessing a Homeless Population: Services Use and Characteristics.” *Psychiatric Services* 67 (12): 1334–39. https://doi.org/10.1176/appi.ps.201500388.

Meyer, Bruce D., Angela Wyse, Alexa Grunwaldt, Carla Medalia, and Derek Wu. 2021. “Learning about Homelessness Using Linked Survey and Administrative Data.” w28861. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w28861.

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

O’Flaherty, Brendan. 2019. “Homelessness Research: A Guide for Economists (and Friends).” *Journal of Housing Economics* 44 (June): 1–25. https://doi.org/10.1016/j.jhe.2019.01.003.

Peressini, Tracy, Lynn McDonald, and J. David Hulchanski. 2010. “Towards a Strategy for Counting the Homeless.” In *Finding Home: Policy Options for Addressing Homelessness in Canada*. Cities Centre Press.

Roncarati, Jill S., Travis P. Baggett, James J. O’Connell, Stephen W. Hwang, E. Francis Cook, Nancy Krieger, and Glorian Sorensen. 2018. “Mortality Among Unsheltered Homeless Adults in Boston, Massachusetts, 2000-2009.” *JAMA Internal Medicine* 178 (9): 1242. https://doi.org/10.1001/jamainternmed.2018.2924.

Rossi, Peter H., James D. Wright, Gene A. Fisher, and Georgianna Willis. 1987. “The Urban Homeless: Estimating Composition and Size.” *Science* 235 (4794): 1336–41.

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

Shlay, Anne B., and Peter H. Rossi. 1992. “Social Science Research and Contemporary Studies of Homelessness.” *Annual Review of Sociology* 18: 129–60.

Smith, Amy Symens, Charles Holmberg, and Marcella Jones-Puthoff. 2012. “The Emergency and Transitional Shelter Population: 2010,” September, 22.

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

U.S. Census Bureau. 2013. “2010 Census Group Quarters Enumeration Assessment Report,” January.

U.S. Department of Commerce. 2011. “Census 2010: Final Report to Congress,” June.

U.S. Department of Education. 2021. "Understanding Doubled Up," December. https://nche.ed.gov/understanding-doubled-up/

U.S. Department of Housing and Urban Development. 2010. “The 2009 Annual Homeless Assessment Report.”

———. 2012. “A Guide to Counting Sheltered Homeless People (Third Revision),” January.

———. 2014. “Point-in-Time Count Methodology Guide.”

———. 2018. “The 2018 Annual Homeless Assessment Report (AHAR) to Congress, Part 1: Point-in-Time Estimates of Homelessness,” December.

U.S. General Accounting Office. 1985. “Homelessness: A Complex Problem and the Federal Response.”

Wagner, Deborah, and Mary Layne. 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. Washington, D.C: U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-01.pdf.

Wolter, Kirk M. 1986. "Some Coverage Error Models for Census Data." *Journal of the American Statistical Association* 81: 338-346.

Wright, James D., and Joel A. Devine. 1992. “Counting the Homeless: The Census Bureau’s ‘S-Night’ in Five U.S. Cities.” *Evaluation Review* 16 (4): 355–64. https://doi.org/10.1177/0193841X9201600401.

Table 1: Homeless Population Estimates

	Unadjusted estimates			**PIT-only population estimates**				**Adjusted estimates**		
Year	**PIT**	**Census**	**ACS**	Safe Haven	Domestic Violence	Voucher-Based	Non-Shelters	**Census**	**ACS**	**Dual systems**
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*	*(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 internal 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 systems estimate is obtained using methods described in Section 7 of the text. Estimate reported here is the midpoint of the range of estimates in that section.

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

	Population in 2010 Census
<u>A: Census and ACS Scope</u>	
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
<u>B: Census Scope Only</u>	
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 DV population is the same in the PIT and the Census.

Table 3: Characteristics of Sheltered Homeless in PIT and ACS (2016)

Source	ACS	PIT
Includes Domestic Violence?	*No*	*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.

Table 4: 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 5: 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 6: Coverage of Los Angeles and Houston HMIS Shelter Users in the 2010 Census

	Los Angeles									
	All Records		Refinement 1: Excluding 3/31 Exits		Refinement 2: Excluding 3/31 Exits and Winter Shelter Program		Refinement 3: Excl 3/31 Exits, WSP, Entries Before 3/1		**Houston**	
Census Status	*Lower*	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*	*Lower*	*Upper*
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	10500		7000		5800		1300		1400	
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	10420		6901		5738		1258		1480	

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 unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table 7: 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	3697		2042		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 unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table 8: 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: Shelter	Housed	Other Census Status	Status Unknown	Total People
			(1)	*(2)*	*(3)*	*(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 .5	*Lower*	0.231	0.170	0.161	0.417	2700
		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 unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table 9: Distribution of GQ Codes for HMIS Shelter Users Appearing in "Other GQ" Types in Census

GQ Code	Category	Los Angeles			Houston
		All records	Refinement 1: Excl 3/31 exits	Refinement 2: Excl 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 unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table 10: HMIS Shelter Users with Multiple Statuses in Census

		Housed +				Sheltered +			Unshel. +	More than two	Other comb-ination	Shelter Users with Duplicate Records	Total Records
		Housed +	Shelt	Unshelt	Other GQ	Sheltered	Unshelt	Other GQ	Unshelt				
Los Angeles	**All records**												
	Lower bound	247	376	131	144	19	37	86	16	91	24	1172	10500
	Upper bound	289	479	188	199	20	57	134	19	139	48	1570	10500
	Refinement 2												
	Lower bound	78	351	61	69	13	34	82	-	79	16	782	5800
	Upper bound	83	448	84	87	14	51	126	-	120	26	1037	5800

		Housed +				Sheltered +			Unshel. +	More than two	Other comb-ination	Shelter Users with Duplicate Records	Total Records
		Housed +	Shelt	Unshelt	Other GQ	Sheltered	Unshelt	Other GQ	Unshelt				
Houston	**All records**												
	Lower bound	22	67	-	19	-	-	-	-	-	31	139	1400
	Upper bound	23	71	-	20	-	-	-	-	-	32	147	1400

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 unauthorized 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 11: Homeless with duplicate housed or other GQ records in Census

	All people		Has at least one housed record			Has at least one other (non-homeless) GQ record		
Homeless type	Number of Records	Unique PIKs	Unique PIKs	Weighted population estimate		Unique PIKs	Weighted population estimate	
				LB	*UB*		*LB*	*UB*
Shelter	209000	143000	26500	43280	49020	1400	2235	3002
Soup Kitchen	162000	67000	29000	72670	84800	1200	2924	4078
Food Van	11500	4900	2300	5588	6399	80	229	305
TNSOL	36500	6300	1900	10660	12830	100	586	835
Homeless type				Share of all records			Share of all records	
				LB	*UB*		*LB*	*UB*
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 unauthorized 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 12: 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	59500	0.939	57000		53500		3500
Age exactly the same	0.709	59500	0.775	53000	0.819	47500	0.099	3100
Age within one year	0.756	59500	0.811	53000	0.855	47500	0.120	3100
Age within five years	0.867	59500	0.903	53000	0.942	47500	0.299	3100
Same race	0.812	59500	0.851	51000	0.862	46000	0.670	2800
Same Hispanic status	0.874	59500	0.890	48000	0.900	43500	0.762	2800
Same state	0.893	59500	0.893	59500	0.893	53500	0.890	3500
Same county	0.806	59500	0.806	59500	0.805	53500	0.808	3500

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 unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-008.

Table 13: Household Characteristics of Homeless Individuals with a Duplicate Housed Record in 2010 Census

Homeless Record Type	Relationship to household head: Household head	Spouse or partner	Child (adult or minor)	Other relative	Other nonrelative	Household type: Lives alone	Lives with family	Lives with non-family	N
Shelter	0.382	0.095	0.318	0.125	0.080	0.185	0.728	0.087	26500
Soup Kitchen	0.516	0.120	0.183	0.100	0.081	0.268	0.616	0.117	29000
Food Van	0.483	0.161	0.197	0.096	0.063	0.198	0.713	0.089	2300
TNSOL	0.386	0.146	0.271	0.111	0.086	0.115	0.790	0.095	1900

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 unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-008.

Table 14: 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	3,368	3,660	2,976	3,212	392	448	2,750	2,966
Unsheltered	1,157	1,294	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	3,101	3,373	1,674	1,778	1,427	1,595	852	902
	Upper bound	*Lower bound*	*Upper bound*	*Lower bound*	*Upper bound*	*Lower bound*	*Upper bound*	*Lower bound*
Status unknown	3,243	2,414	1,400	957	1,843	1,457	338	157
Total	11,899	11,899	7,218	7,218	4,681	4,681	4,521	4,648

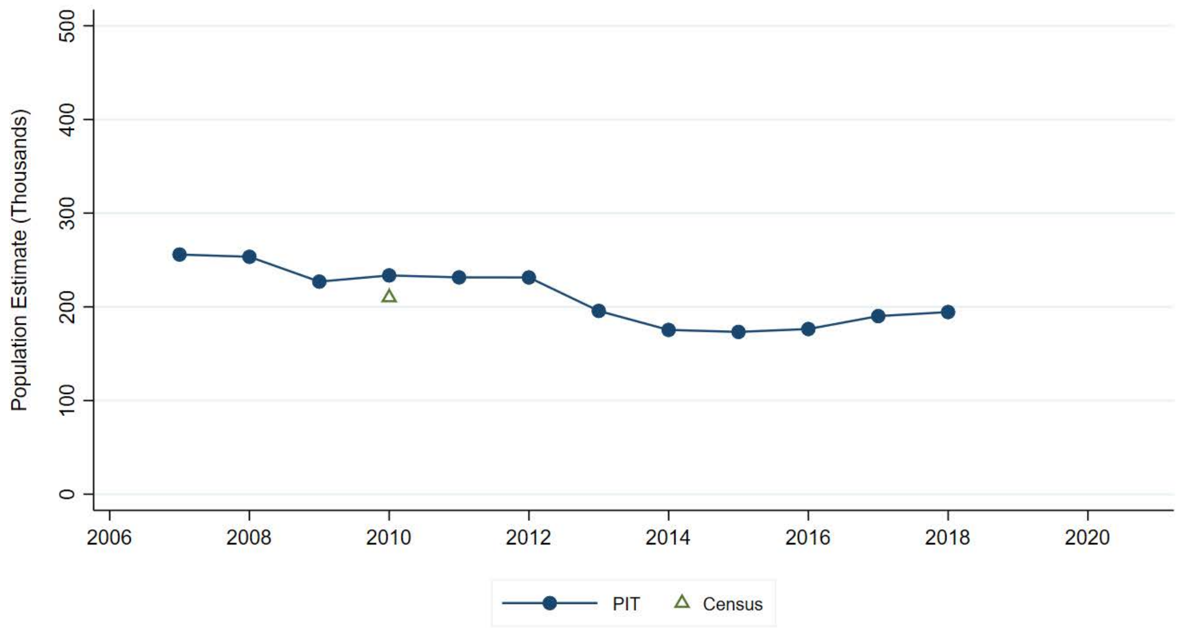
B: Weighted shares

	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	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Source: 2010 Census, 2004-2014 Los Angeles HMIS datasets, 2004-2015 Houston HMIS datasets

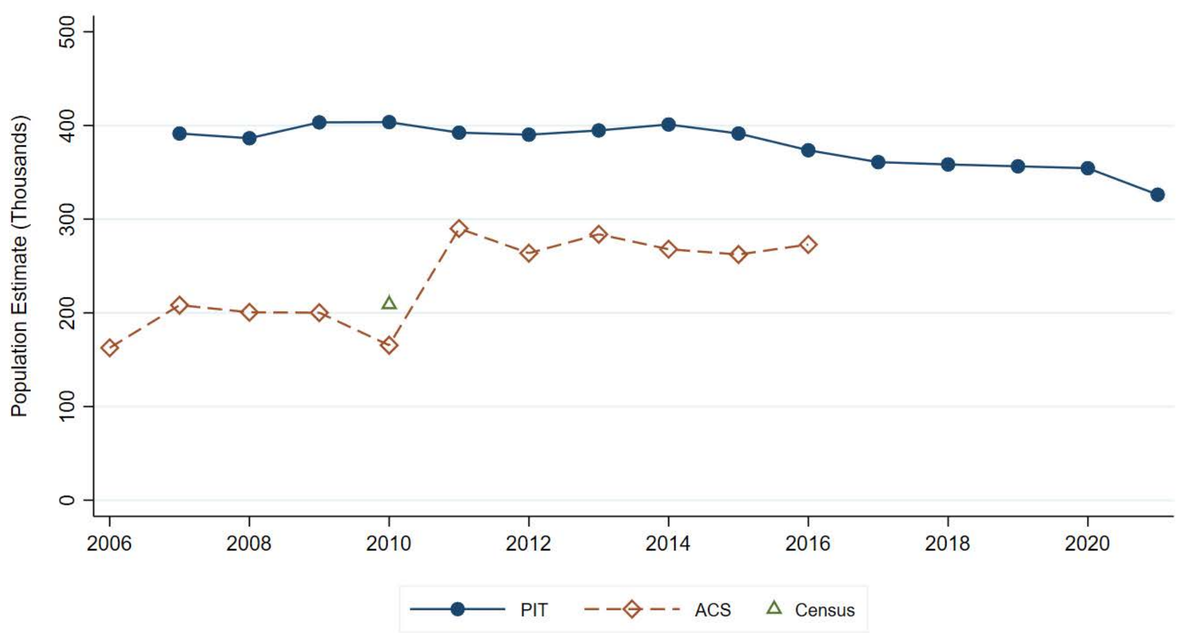
Notes: Table indicates weighted counts in Census statuses, calculated as the sum of weighted totals from Los Angeles and Houston HMIS datasets. Columns (2) indicate bounds on the sum of Houston and L.A. weighted totals under Refinement 2. Columns (3) indicate bounds on the the difference between (1) 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).

Figure 1. Unsheltered homeless population estimates in the PIT and Census



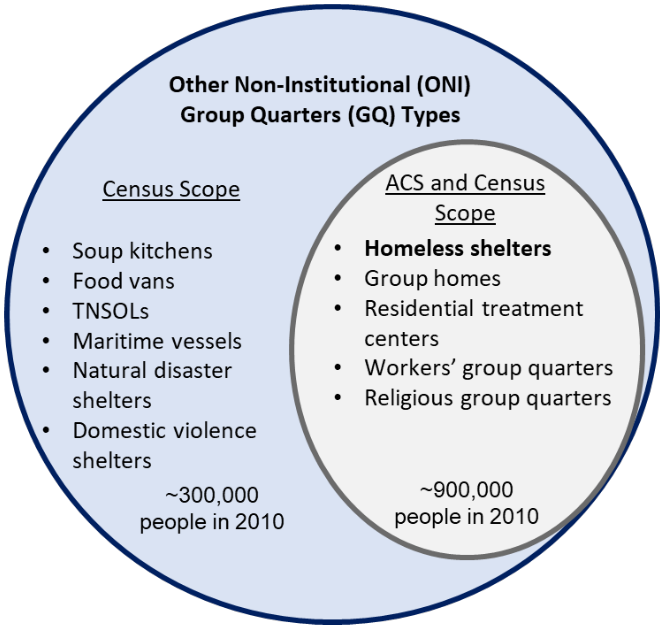
Sources: 2010 Census, 2007-2018 PIT files
Approved for release by the Census Bureau's Disclosure Review Board, authorization number CBDRB-FY2022-CES005-015

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



Sources: 2010 Census, 2006-2018 ACS, 2007-2018 PIT files
Approved for release by the Census Bureau's Disclosure Review Board, authorization number CBDRB-FY2022-CES005-015

Figure 3. Graphical representation of Other Non-Institutional (ONI) group quarters types



Note: circles not to scale.

Figure 4. Sheltered homeless population estimates in the PIT, ACS, and Census with definitional and weighting adjustments and dual systems estimate

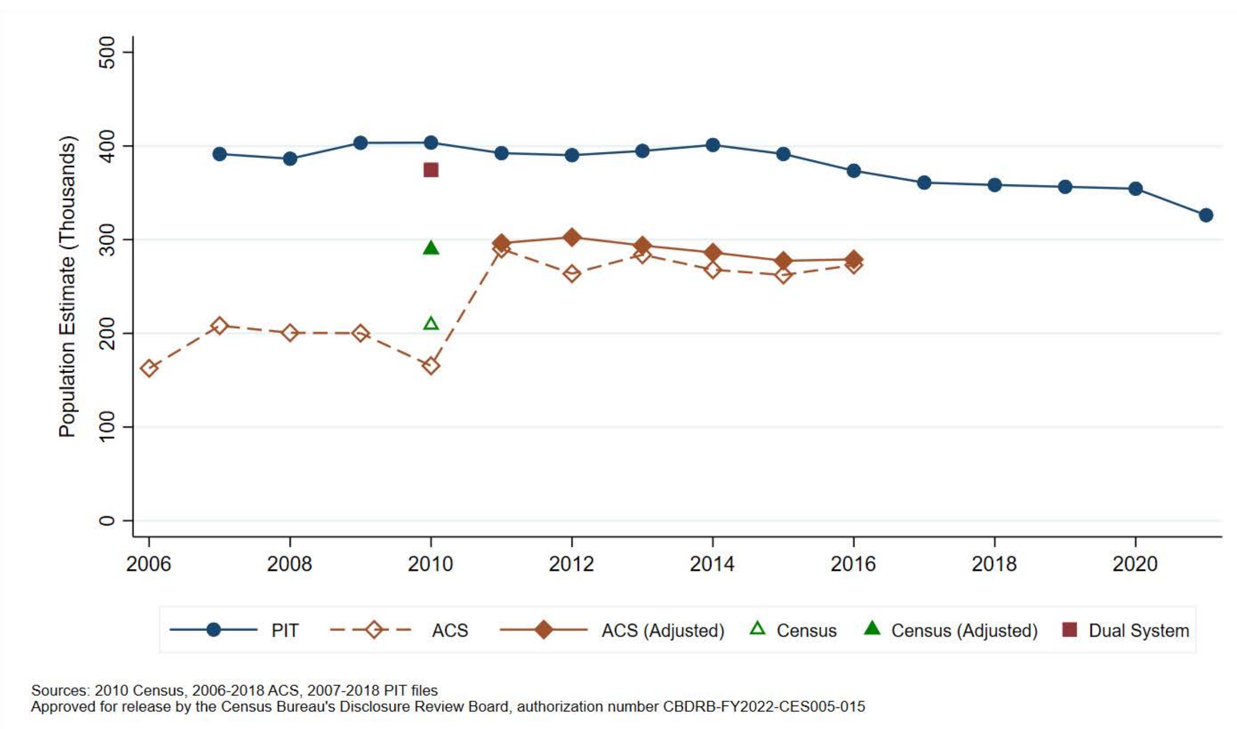


Figure 5. Los Angeles HMIS Data Quality

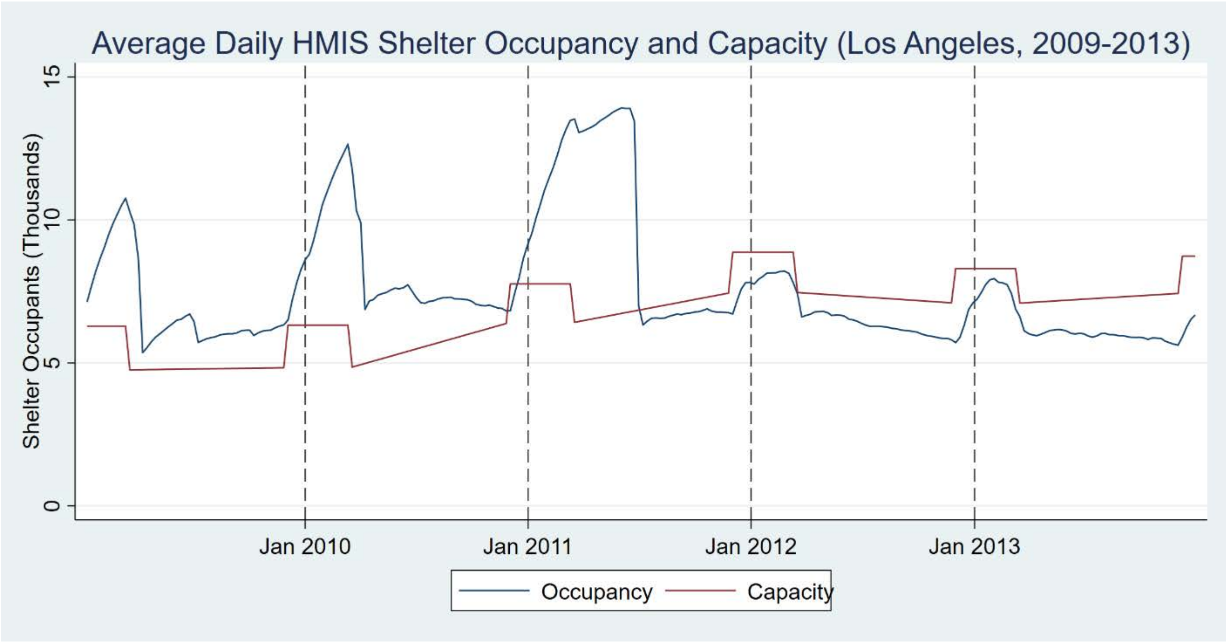
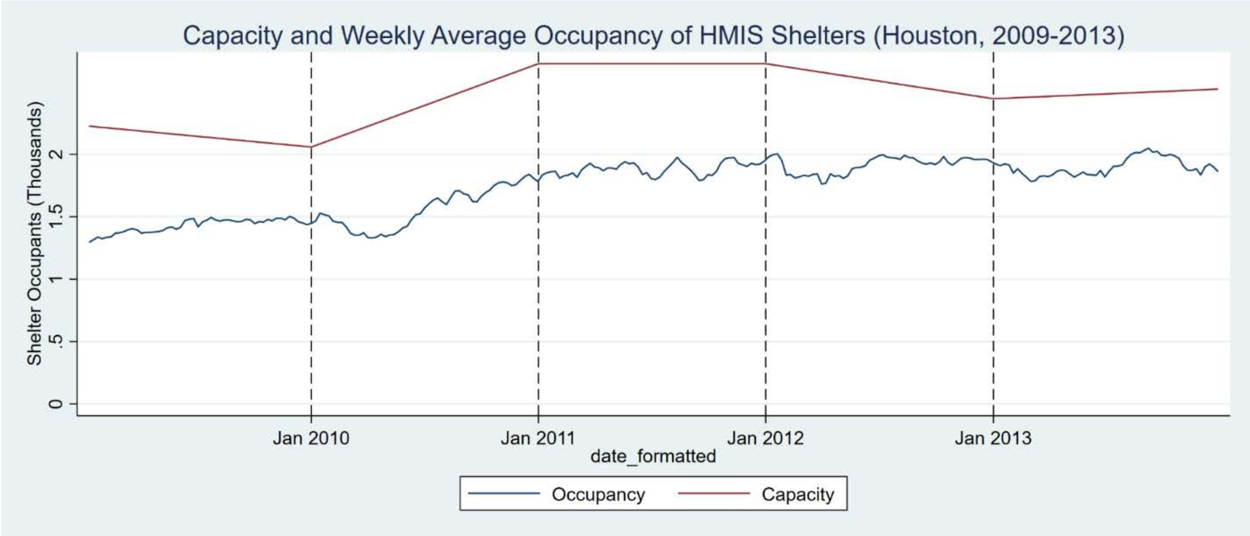


Figure 6. Houston HMIS Data Quality



Appendix

A1. Possible sources of imprecision in estimates

In addition to the estimation of reasons for CoC-level differences in the relative magnitude of estimates described in the text, we consider here potential sources of imprecision in CoC-level homeless population estimates, namely:

- *HMIS participation rate.* We expect CoCs with a larger share of beds tracked through HMIS to have more precise sheltered PIT estimates, since many CoCs extrapolate from HMIS records to obtain their sheltered estimate.
- *New (versus carried-over) PIT.* We expect CoCs that conducted their PIT in 2010 instead of carrying over a previous year's count to have more precise PIT estimates.
- *Year-to-year variability in PIT estimate.* Large year-to-year fluctuations in a CoC's sheltered or unsheltered PIT estimate in years after 2010 may reflect a lower quality CoC methods or technical capacity in 2010.
- *CoC population size.* Population size could be associated with more precision if idiosyncratic factors average out in more populous CoCs. Alternatively, it could be more difficult to administer the counting operation in more populous CoCs, leading to less precise estimates.

We once again incorporate a vector of other CoC characteristics into our model, although again it is not clear that we should expect a particular pattern of relationships between these variables and imprecision.

A1.1 Modeling imprecision in estimates

We define imprecision as the absolute percentage difference between the PIT and Census count and use OLS to estimate the following models: [38]

[38] We again use Weighted Least Squares (WLS) to account for the likely presence of heteroskedasticity in our variables relative to CoC population size.

$$\frac{|S_i^{Cen} - S_i^{PIT}|}{\frac{1}{2}(S_i^{Cen} + S_i^{PIT})} = \alpha_0 + \alpha_1 HMIS_i + \alpha_2 2010\ PIT_i^S + \alpha_3 Var\ PIT_i^S + \alpha_4 \ln(Pop_i) + Char_i' \delta + \epsilon_{3,i}$$

$$\frac{|U_i^{Cen} - U_i^{PIT}|}{\frac{1}{2}(U_i^{Cen} + U_i^{PIT})} = \delta_0 + \delta_1 HMIS_i + \delta_2 2010\ PIT_i^U + \delta_3 Var\ PIT_i^U + \delta_4 \ln(Pop_i) + Char_i' \delta_5 + \epsilon_{4,i}$$

where S_i (U_i) is CoC i's sheltered (unsheltered) homeless count, $HMIS_i$ is the CoC's HMIS participation rate, $2010\ PIT_i^S$ ($2010\ PIT_i^U$) is an indicator for whether a CoC conducted a sheltered (unsheltered) PIT in 2010 instead of carrying over a previous year's estimate, $Var\ PIT_i^S$ ($Var\ PIT_i^U$) is an indicator for whether a CoC's sheltered (unsheltered) PIT ever fluctuated by more than 0.5 log points year-to-year from 2011-2012 onwards, and $\ln(Pop)_i$ is the log of the CoC's population.

We estimate our models on both a limited sample (keeping only CoCs where the minimum of the sheltered PIT and Census was greater than 100 and where the minimum of the unsheltered PIT and Census was greater than 50) and on a full sample, which includes all CoCs. While less informative on sources of imprecision affecting national trends, the full sample estimation tells us about the precision of estimates in all CoCs, including small ones.

A1.2 Results

We observe that CoCs with greater HMIS participation, i.e., a greater share of beds tracked through the CoC's HMIS database, tend to have more precise sheltered homeless estimates, as predicted, although the magnitude of this effect is small.[39] Increasing HMIS participation by 10 percentage points is associated with a 0.015 unit decrease in imprecision, relative to a mean of 0.383.

Two other aspects of PIT methodology – whether a sheltered PIT was conducted in 2010 or carried over from 2009, and whether the unsheltered PIT tended to vary substantially year-to-year from 2012 onwards – were significant with the expected sign, but only when considering the full sample of CoCs, not just those where the minimum of the two counts exceeded a certain threshold. These aspects of PIT methodology may only matter for small CoCs.

[39] Table A3 in the appendix displays key results from these imprecision estimations.

Finally, we do find that CoC characteristics are jointly significant at the five percent level in some of our models, but the patterns displayed by these coefficients do not lend themselves readily to hypotheses about the mechanism underlying these associations.

A2. Additional results from microdata comparisons

A2.1 Characteristics of recent HMIS shelter occupants missed by the Census

In Table 7, we saw that L.A. 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 2,500 prior to refinements to fewer than 1,000 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 A12. We observe that dropped individuals – i.e., 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 shelter 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 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. 2021).

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.

A2.2 Examining the 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 A13 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.[40]

[40] 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.

Appendix Tables and Figures

Table A1: Results on Relative Magnitude of Sheltered Estimates

	Census/PIT Sheltered Estimates
Census/PIT Shelter Count Ratio	0.395***
	(0.0596)
Characteristics (Urban, Poor, Density, Rent)	
p-value from test of joint insignificance	0.2933
Weather (Day-of and Month-of Temp and Prcp)	
p-value from test of joint insignificance	0.0704
Intercept	0.456***
	(0.162)
N	328
R-squared	0.241
F-statistic	5.922
p-value	7.91e-10

White standard errors in parentheses; weighted for heteroskedasticity.
* p<0.10 ** p<0.05 *** p<0.01

Table A2: Results on Relative Magnitude of Unsheltered Estimates

Census/PIT Unsheltered Estimates - Marginal Effect at Mean		
	(1)	*(2)*
Weather Model:	Level Terms	Level and Sq Terms
Day-of-Count Weather		
Census Min Temp (F)	0.0177	0.0409
Prediction: (+)	(0.0401)	(0.0403)
PIT Day Temp (F)	0.0299	0.0152
Prediction: (+)	(0.0207)	(0.0207)
Census Total Precip (mm)	0.0398	0.0093
Prediction: (-)	(0.0267)	(0.0321)
PIT Total Precip (mm)	0.0385*	0.0334
Prediction: (+)	(0.0220)	(0.0612)
Month-of-Count Weather		
Census Min Temp (F)	0.0213	0.0102
Prediction: (+)	(0.0509)	(0.0622)
PIT Min Temp (F)	-0.120***	-0.1223***
Prediction: (-)	(0.0291)	(0.0314)
Census Avg Daily Precip (mm)	0.0661	0.2728**
Prediction: (-)	(0.0906)	(0.1252)
PIT Avg Daily Precip (mm)	-0.157***	-0.2610**
Prediction: (+)	(0.0591)	(0.1310)
CoC Characteristics		
p-value from test of joint insignif.	0.1092	0.2080
Intercept	2.484	7.595***
	(1.60)	(2.26)
N	239	239
R-squared	0.281	0.33
F-statistic	6.169	6.341
p-value	2.29E-09	2.51E-13

White standard errors in parentheses; weighted for heteroskedasticity.

* p<0.10 ** p<0.05 *** p<0.01

Notes: All estimates are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A3: Results from Imprecision Regressions

	Sheltered Abs Percent Difference		Unsheltered Abs Percent Difference	
	Limited Sample	*All CoCs*	*Limited Sample*	*All CoCs*
	(1)	(2)	(3)	(4)
Ln Population	-0.0100	-0.0423**	-0.00987	-0.0702**
	(0.0168)	(0.0174)	(0.0364)	(0.0311)
HMIS Participation	-0.146*	-0.186**	-0.107	-0.141
Sheltered prediction: (-)	(0.0836)	(0.0864)	(0.156)	(0.133)
Sheltered PIT in 2010	-0.0468	-0.0917**		
Sheltered prediction: (-)	(0.0438)	(0.0444)		
Variable Sheltered PIT	0.0236	0.0210		
Sheltered prediction: (+)	(0.0725)	(0.0706)		
Unsheltered PIT in 2010			0.103	0.0175
Unsheltered prediction: (-)			(0.0677)	(0.0622)
Variable Unsheltered PIT			0.00833	0.185***
Unsheltered prediction: (+)			(0.0696)	(0.0615)
CoC Characteristics				
p-value from test of joint insignif.	0.0174	0.0812	0.2648	0.0003
Intercept	0.488*	1.264***	1.156*	2.892***
	(0.264)	(0.272)	(0.597)	(0.486)
N	328	385	239	385
R-squared	0.0618	0.0817	0.0254	0.104
F-statistic	2.393	3.639	1.118	6.679
p-value	0.0162	0.000428	0.352	3.49e-08

White standard errors in parentheses; regressions weighted for heteroskedasticity.

* p<0.10 ** p<0.05 *** p<0.01

Notes: All estimates are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A4: Predicted Effect of Setting Census Day Weather Equal to PIT Day Weather on Census Sheltered Estimate and Mean Census/PIT Ratio

	(1)	*(2)*
Weather Model:	Level Terms	Level and Sq Terms
Predicted Effect on Aggregate Census Sheltered Estimate		
Actual Census Estimate	279,000	279,000
Predicted Census Estimate with PIT Weather	313,800	311,600
Predicted Percent Change	12.47%	11.68%
Predicted Effect on Mean Census/PIT Ratio		
Actual Mean Census/PIT Ratio	0.765	0.766
Predicted Mean Census/PIT Ratio with PIT Weather	0.847	0.846
Predicted Percent Change	10.62%	10.49%

Sources: 2010 PIT, 2010 Census

Notes: Table displays the predicted sheltered Census estimate and the predicted mean Census/PIT ratio that would occur if we set the temperature and precipitation on the day of and in the month of the Census equal to that of the PIT. We include only the subset of CoCs where the minimum of the PIT and Census sheltered count was greater than or equal to 100. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A5: Predicted Effect of Setting Census Day Weather Equal to PIT Day Weather on Census Unsheltered Estimate and Mean Census/PIT Ratio

	(1)	*(2)*
Weather Model:	Level Terms	Level and Sq Terms
Predicted Effect on Aggregate Census Unsheltered Estimate		
Actual Census Estimate	188,000	188,000
Predicted Census Estimate with PIT Weather	173,000	198,500
Predicted Percent Change	-7.98%	5.59%
Predicted Effect on Mean Census/PIT Ratio		
Actual Mean Census/PIT Ratio	2.226	2.347
Predicted Mean Census/PIT Ratio with PIT Weather	1.448	2.138
Predicted Percent Change	-34.95%	-8.90%

Sources: 2010 PIT, 2010 Census

Notes: Table displays the predicted unsheltered Census estimate and the predicted mean Census/PIT ratio that would occur if we set the temperature and precipitation on the day of and in the month of the Census equal to that of the PIT. We include only the subset of CoCs where the minimum of the PIT and Census unsheltered count was greater than or equal to 50. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A6: Marginal Effect of Weather Terms at Mean with Standard Errors (Models with Squared Weather Terms)

	Sheltered	Unsheltered
Day-of-Count Weather		
Census Day Min Temp	-0.0007	0.0409
	(0.0030)	(0.0403)
PIT Day Min Temp	0.0028	0.0152
	(0.0027)	(0.0207)
Census Day Total Precip	-0.0042	0.0093
	(0.0031)	(0.0321)
PIT Day Total Precip	-0.0097	0.0334
	(0.0056)	(0.0612)
Month-of-Count Weather		
Census Month Avg Temp	-0.0031	0.0102
	(0.0078)	(0.0622)
PIT Month Avg Temp	-0.0014	-0.1223***
	(0.0038)	(0.0314)
Census Month Avg Daily Precip	0.0271	0.2728**
	(0.0175)	(0.1252)
PIT Month Avg Daily Precip	-0.0157	-0.2610**
	(0.0155)	(0.1310)

Sources: 2010 PIT, 2010 Census
Notes: Table displays the marginal effect of weather and climate variables at their means in the model that includes both level and squared seasonality terms. Standard error and significance of the linear combination of level and squared term is reported. All estimates are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A7: Bias Regressions - Hypothesis Tests

Test Description:	Joint significance of level weather terms	Joint significance of squared weather terms	Weather variables constrained to differences	Joint significance of CoC characteristics	Joint significance of CoC characteristics
Base Model:	Level Weather	Squared Weather	Level Weather	Level Weather	Squared Weather
Sheltered					
F-statistic	1.8330	0.3844	3.2730	1.2410	0.6660
p-value	0.0704	0.9286	0.0119	0.2933	0.6160
Unsheltered					
F-statistic	7.3660	3.5330	5.9020	1.9130	1.4840
p-value	0.0000	0.0007	0.0002	0.1092	0.2080

Sources: 2010 PIT, 2010 Census
Notes: All estimates are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A8: Imprecision Regressions - Hypothesis Tests

Test:	Joint significance of CoC characteristics			
	Sheltered		Unsheltered	
Sample:	Limited	Full	Limited	Full
F-statistic	3.0440	2.0920	1.3160	5.5030
p-value	0.0175	0.0812	0.2648	0.0003

Sources: 2010 PIT, 2010 Census
Notes: All estimates are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A9: Summary Statistics for CoC Regressions

	Sheltered		Unsheltered	
	Mean	SD	Mean	SD
Outcomes				
Census (Adjusted)	850.80	2343.00	777.20	1215.00
PIT	1192.00	3026.00	901.00	1815.00
Census/PIT	0.77	0.29	2.27	2.81
Abs Percent Difference	0.38	0.28	0.79	0.51
Number of Shelter Addresses				
Census Shelter Addresses (Adj, Address Correction)	27.93	36.39		
PIT Shelter Addresses	36.21	44.44		
Shelter Address Ratio (Census/PIT)	0.83	0.32		
CoC Population Characteristics				
CoC Population (Millions)	0.89	1.21	1.04	1.36
Pop Share in Poverty	0.14	0.05	0.15	0.05
Pop Share in Urban Area	0.84	0.17	0.85	0.17
Pop Density (1000 people per km^2)	0.50	1.08	0.51	1.16
Median Rent Index	1.04	0.24	1.06	0.25
Accuracy Variables				
HMIS Participation	0.78	0.21		
PIT in 2010	0.78	0.41	0.59	0.49
Variable PIT	0.05	0.22	0.42	0.49
Day Weather				
Census Temperature (F)	39.70	7.78	40.53	8.21
PIT Temperature (F)	27.17	14.10	29.65	14.67
Census Precipitation (total, mm)	7.32	15.45	6.52	14.42
PIT Precipitation (total, mm)	1.75	5.04	2.37	6.15
Month Weather				
Census Temperature (F)	47.01	7.25	48.16	7.90
PIT Temperature (F)	34.33	12.58	37.42	13.57
Census Precipitation (avg daily, mm)	3.52	2.63	3.38	2.57
PIT Precipitation (avg daily, mm)	2.34	1.74	2.28	1.77
Number of CoCs				

Source: 2010 PIT, 2010 Census, 2010 ACS, 2008-2016 HIC
Notes: Census counts are adjusted to include PIT-only population (domestic violence shelters, etc.). Census shelter count is adjusted for the PIT-only population and PIT shelter count is adjusted for the presence of multisite shelters and duplicate addresses. All counts and shares are rounded per Census Bureau disclosure rules. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A10: 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 unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A11: 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		10900	

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.

Table A11: Probability of L.A. HMIS Shelter Entry and Hazard Rate for Exit

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Entry Probability												
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 unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A12: 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 unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-006.

Table A13: 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*	*Upper*	*Lower*	*Upper*
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 unauthorized disclosure of confidential information and has approved the disclosure avoidance practices applied to this release, authorization number CBDRB-FY2022-CES005-008.