Counting the Homeless:
Improving Knowledge of the Unsheltered Homeless Population

Christopher Weare, Ph.D.
Goldman School of Public Policy
University of California, Berkeley
weare@berkeley.edu

December 2019

“If you can’t measure it, you can’t improve it.”
Peter Drucker

It is time to rethink and improve Point-in-Time (PIT) counts -- yearly, one-night counts of individuals experiencing homelessness. For communities that already conduct high-quality PIT counts of the population of unsheltered homeless individuals, the method proposed here provides a valuable statistical check of their accuracy.[1] For the larger number of communities that lack the resources and capacity to conduct high-quality counts, these methods provide an avenue for them to produce more accurate counts and to improve the analysis of trends over time.

PIT counts conducted since 2005 constitute a major advance in efforts to understand homelessness and play a central role in the development of policy. In 1989, when Peter Rossi published his seminal book, *Down and Out in America: The Origins of Homelessness*, national estimates of the number of people experiencing homelessness differed by an order of magnitude, ranging between 250,000 to 2.2 million. [1] This wide range resulted from a mix of ad hoc methods that lacked a systematic, statistically valid approach. In his book, Rossi proposed generating more rigorous estimates based on a random sample of geographic areas and conducted a series of such pathbreaking surveys in Chicago.

Rossi’s work inspired the adoption of new methods for PIT counts. They now entail a nation-wide effort that is coordinated by the Department of Housing and Urban Development (HUD) which promulgates guidelines on appropriate count methods. Communities throughout the United States design and administer counts employing a variety of methods. They strive to adhere to HUD guidelines, but they are constrained by resource limitations including access to volunteer enumerators and research expertise.

These counts provide estimates of the size of the homeless population, of trends in homelessness, and of the composition of the homeless subpopulations including veterans, youth, people suffering from mental illness, families and the chronically homeless. Their results are consolidated into a report delivered yearly to Congress, drive policy discussions, determine the allocation of resources, and receive extensive media attention.

PIT counts, nevertheless, have been broadly criticized for undercounting the overall number of people experiencing homelessness, specifically missing certain subpopulations and other inconsistencies. [2-5] Philip Mangano, the former head of the U.S. Interagency Council on Homelessness has called PIT counts “one of the most unscientific activities that determines policies ever derived by the federal government.” [6] Also, because of the significant resources required to conduct a PIT count, they are only conducted annually or biannually, meaning that

[1] PIT counts include two main components a sheltered and an unsheltered count. The sheltered count uses HMIS data and surveys to count individuals housed in shelters and transitional housing programs. This brief is focused on the unsheltered count.

data quickly becomes outdated and that PIT counts provide limited insight into the seasonality of homeless and the incidence of short-term or episodic homelessness. Despite these known flaws, the methods employed to conduct these counts have changed little since 2005.

This policy brief proposes a major improvement to PIT counts based on an estimation technique called the multiple-list method[2]. It entails combining PIT counts with information from Homeless Management Information Systems (HMISs), which track the services provided by a homelessness crisis system such as emergency shelter bed nights and street outreach interactions. Unlike PIT count, the HMIS data has been continuously improved. Data standards have been updated 5 times in the last 15 years and HUD has emphasized the use of these data to examine system performance.[3]

The quality and availability of HMIS data have improved to the point that they offer an important source from which population estimates can be calculated. Specifically, communities that engage in street outreach programs, can combine these data with information obtained from the traditional PIT count to produce higher quality estimates. This multiple-list method offers several advantages for communities and national policy makers. It can be implemented inexpensively in communities that support on-going street outreach programs because the necessary data is already available. It can estimate the size of the entire population of people experiencing homelessness accurately with known confidence intervals. While the two-list estimates entail sampling error, given the resources currently available for PIT counts, the two-list method provides more accurate estimates compared to other feasible estimation methods. (See Appendix A for details).

PIT counts constitute a nation-wide effort involving multiple stakeholders including HUD, the nearly 500 continuums of care, and advice provided by consulting firms and academics. To benefit from the multiple-list method does not require wholesale changes in national policies. A single Continuum of Care could use the method independently. Nevertheless, to maximize the benefits of better estimates would require a range of actions by stakeholders. They include:

1. HUD should include the multiple-list method in its PIT count guidance as an acceptable method for community PIT counts. HUD can also provide guidance on how to implement the method.
2. Many communities already conduct high-quality PIT counts based on either a census or a sample of geographical areas that is then extrapolated to the Continuum's full territory.[4] For them the multiple-list estimate should be employed as a complement to the current count. It provides a statistical check on the existing count and can identify sources of error.

[2] These methods are also referred to as multiple system estimates, capture and recapture methods, or mark and capture methods. The latter two names hark back to the roots of these estimators in the population ecology field.
[3] For example, the System Performance Metrics employed by HUD are based on HMIS data, and HUD is in the process of rolling out a suite of services called Stella that provides dashboard visualizations of system performance.
[4] HUD PIT guidelines advises communities to employ these methods, but it acknowledges that many CoCs do not have access to the expertise and other resources necessary to implement them. [7]

3. Many communities rely on less rigorous methods due to resource constraints. They should consider employing the multiple-list method to improve the quality of their PIT estimates and make estimates more comparable from year to year.
4. The consulting and academic communities that advise Continuums of Care should offer a two-list estimate as either a complement or substitute to existing methods.

This policy brief first reviews the limitations of current PIT counts. It then introduced the multiple-list method that combined PIT and HMIS data. It briefly reviews possible extensions of these methods to provide communities with more frequent estimates of the dynamics of their homeless populations and it ends with policy recommendations.

Limitations of PIT Counts

HUD encourages Continuums of Care (CoCs), the local entities that conduct PIT counts, to employ survey methods that validly estimate the entire population experiencing homelessness in the CoC's geographic area. As stated in its most recent Notice:

> CoCs must ensure that their count estimate accurately reflects what they believe to be the entire sheltered and unsheltered population for the CoC's entire geographic area. For example, if a CoC only counts unsheltered people in selected areas, they need to consider whether there are likely unsheltered homeless persons in other areas of the CoC and, if so, how to account for them. This is particularly important when entire counties, communities, or larger geographic areas are not covered. ***CoCs should use sampling and extrapolation methods to account for areas that were not included in the unsheltered count if there is any possibility an unsheltered person could be found there.*** [emphasis added]. [7]

Many CoCs strive to adhere to these guidelines. In a review of 2017 California PIT counts eight of 43 Californian CoCs claimed to cover their entire geographic area.[5] Nevertheless, there are numerous reasons to believe that even these counts miss a significant number of people experiencing homelessness. First, there are misidentification issues when PIT counters fail to recognize that a person who on the streets during the count is in fact homeless. Researchers tested this possibility by placing plants on the streets during a PIT count in New York City. [3] They found that almost 30% of their plants were missed by PIT counters. Second, PIT counts focus on counts of people on the streets. Many individuals living in unsheltered places, nevertheless, seek to remain hidden and stay in abandoned buildings and other places where PIT counters do not enter. A study in Los Angeles investigated this possibility with a post-PIT household telephone survey. [8] It inquired whether the respondents were aware of any homeless individuals sleeping on private property in their neighborhood. Based on the survey, they estimated that including these hidden homeless in the PIT would increase the unsheltered

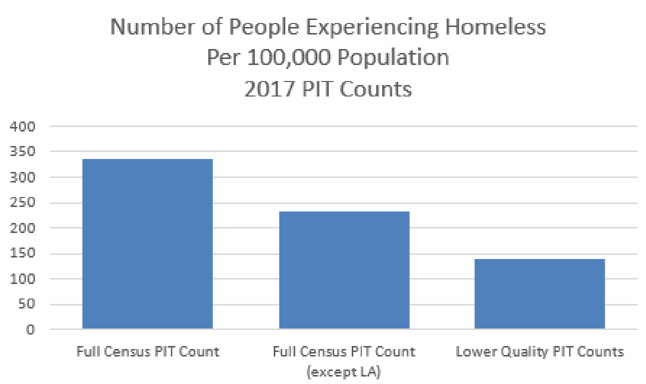
[5] This count is a rough estimate. PIT Count reports only provide cursory reviews of their survey methods, requiring some educated guesses on the actual methods employed. Interestingly, all the CoCs that cover their entire area claim to conduct a census of all areas. No CoC took a sample of areas and extrapolation to their entire region.

count by over 50%. Third, many people included in the count are sleeping in tents or vehicles so the count must estimate how many individuals are in each.

There are also concerns about the quality of the count. PIT counts are major community efforts that rely on the contributions of hundreds if not thousands of volunteers. These volunteers receive limited training and reports in the news suggest that inconsistent counting methods can lead to both over and undercounts by volunteers. For example, CoCs commonly report that an increase in the number of volunteers leads to an increase in the number of individuals identified in comparison to the previous PIT count. Nevertheless, if a CoC is properly employing the census or sampling and extrapolation methods recommended by HUD, the number of volunteers should have no effect on the estimated totals.[6]

The majority of CoCs do not employ the best methods recommended by HUD. Rather than trying to account for its entire geography, many CoCs rely on a known locations method that focuses the PIT count on areas where community members report seeing encampments prior to the PIT night. Alternatively, some CoCs rely on a service-based count that interviews individuals at soup kitchens, day centers, and other service providers and asks whether they were unsheltered on the PIT night.

Figure 1



These less thorough methods lead to significant undercounts compared to CoCs that strive to cover their full geography. The population of a CoC explains most (about 83%) of the differences in the number of homelessness in different California CoCs. Thus, comparing the number of homeless per capita is a useful first-cut comparison of the incidence of homelessness. As seen in Figure 1, the per capita rates of homelessness in CoCs that make efforts to count all homeless in their area are almost 2 ½ times higher than in CoCs that report using less thorough methods.[7] Even when one excludes Los Angeles, which has by far the largest homeless population in the State, CoCs with higher quality PIT counts report 67% more homeless.

More evidence comes from Sacramento's 2019 PIT count. [9] It engaged in an experiment in which it expanded its count beyond the known locations on which it had previously focused. It sent count teams to 64 randomly selected areas where there had been no reports of homeless encampments. Counters identified 286 people in those areas providing conclusive evidence that PIT counts that ignore large portions of a CoC's geography miss numerous people experiencing homelessness.

[6] A larger number of volunteers would improve the accuracy of the estimate because sampling a greater number of areas narrows the confidence interval of the estimate, but the estimates should only differ by sampling error.
[7] The CoCs that report conducting a full census of geographic areas in 2017 are Los Angeles, San Francisco, San Diego, Alameda County, Santa Clara County, San Mateo County, Santa Barbara County, and Pasadena

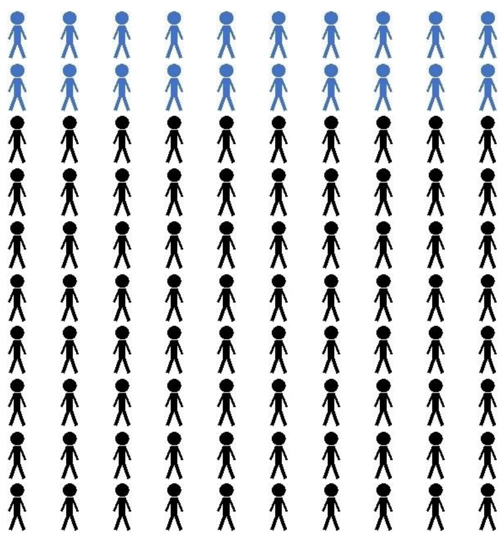
Because resources from both the Federal and State governments are allocated based on PIT counts, the methodological differences that distort PIT counts have significant impacts. For example, in 2019 California will allocate $650 million to localities. If all California CoCs had conducted high-quality PIT counts, there would be likely have been a shift of tens of millions of dollars away from the large coastal areas that are more likely to conduct full census counts to inland areas that typically employ lower-quality estimation methods that lead to undercounts.

In sum, current methods for counting the homeless are flawed. They lead to dramatic undercounts of the total, and because CoCs employ different methods, some of which are less likely to capture all of the homeless, the counts distort the allocation of program funding.

The Multiple-List Method

Alternative approaches to estimating the size of homelessness populations exist and can fruitfully complement and potentially replace current methods. This family of approaches is called the multiple-list method. It relies on extracting two or more samples from an unknown population and estimates the size of the total population based on the overlap between samples. The method was pioneered and elaborated by ecologists interested in studying the size of wildlife populations. [10] Nevertheless, it has been employed to estimate the size of homeless populations as well as other hard to count populations such as IV drug users. [11-13] In fact, at the same time that Rossi was conducting his Chicago study, statisticians conducted a two-list estimate of the homeless population in Baltimore, though their work received far less attention. [13]

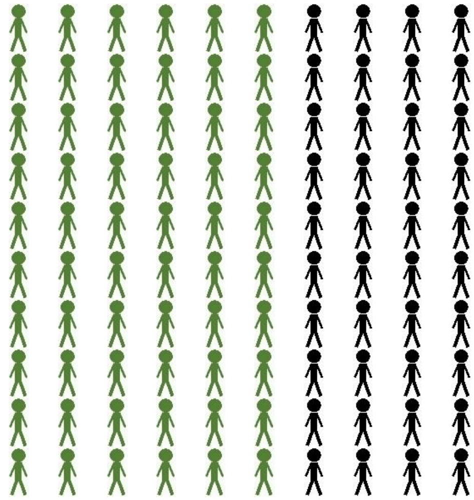
Figure 2: 100 Unsheltered Homeless with 20 (Blue) Enrolled in Street Outreach



A quick example illustrates how the multiple-list method works. Assume that a community, depicted in Figure 2, has 100 individuals who are homeless and unsheltered. Researchers do not know this number but assuming that it is known simplifies the example. Say, the CoC operates a street outreach program that is working with 20 unsheltered clients, colored blue in Figure 2. So, the CoC knows that at least 20 individuals are experiencing unsheltered homelessness but does not know the size of the total population.

Then, on the night of the PIT, counters identify 60 people who are unsheltered, colored green in Figure 3. Researchers then calculate the overlap between these two lists. In this example, it is twelve individuals, colored yellow in Figure 4.

Figure 3: 100 Unsheltered Homeless with 60 Individuals Identified in PIT Count



This overlap allows one to estimate the size of the total population. Street outreach enrollees comprise 20% of the sample collected in the PIT count (12/60 = 20%). This percentage provides a good estimate of the proportion of the total homeless population comprised of the street outreach clients. In other words, if the street outreach clients are 20% of all unsheltered homeless, one would expect that about 20% of a random sample from that population (e.g. the sample collected in the PIT count) should be comprised of street outreach clients. The total population, P, multiplied by 20% equals 20, the number of street outreach clients. Then to estimate P researchers need only divide the number of street outreach clients by 20% to arrive at an estimate of 100 individuals.

$$P \cdot .20 = 20$$

$$P = \frac{20}{.20} = 100$$

For this multiple-list method to work correctly a few assumptions must be met. They are:

1. The population is stable, meaning that no individual experiencing homeless either leaves or enters the area between the time each list is collected.
2. The probabilities of being included on different lists are independent, meaning that being on one list does not either increase or decrease the likelihood that a person will be also be included in other lists.
3. The probability of being included on a list is the same for all homeless individuals.

Figure 4: Overlap (Yellow) between Street Outreach (Yellow and Blue) and PIT Count (Yellow and Green)

Previous efforts to calculate multiple-list estimates of homeless populations faced difficulties with satisfying these assumptions, which has limited their adoption. Early applications of the method relied on client lists from service providers like soup kitchens and emergency shelters. While these lists were readily available, they typically violated basic assumptions. If there was a long gap between the time at which different lists were recorded, it is possible that individuals either entered or left the state of homelessness, violating the stable population assumption. If people who received services from one service provider were more likely to engage in other services, the independence assumption was violated. Finally, the location and targeting of programs like soup kitchens and day service centers could mean that not all people were equally likely to receive services and be on that particular service provider list. All of these violations biased population estimates. While there are statistical models for estimating and correcting for these interdependencies, little work has been done to develop these models for populations of people experiencing homelessness. [14, 15]

The advent of robust street outreach programs provides a solution to these problems.[8] CoCs that maintain an active list of people with whom outreach workers are in contact can compare that list to a list of people who are interviewed by PIT counters. The demographic survey administered by PIT counters need only include a question that requests identifying information to match survey takers with the list of street outreach clients. Typically, the first three letters of the person's first and last name and the month and day of their birth suffices.

Using lists based on street outreach contacts and PIT counts comply with the assumptions for multiple-list estimates to a much greater degree. First, street outreach workers can be asked to verify that their clients were unsheltered on the night of the PIT count. Thus, both lists are taken from the same population on the same night, avoiding the possibility of exits or entries. Second, because street outreach and the PIT are separate activities, the probabilities of being included in each list are independent. The one administrative check to ensure this independence would be that on the night of the PIT count a CoC should avoid sending street outreach caseworkers to the areas in which they normally work where they may be more likely to know the location and recognize current clients.

The one assumption that is more difficult to meet is the assumption that every individual has an equal probability of being included on each list. Studies of PIT counts have shown that certain populations (e.g. youth and others who do not wish to be found) are less likely to be included in a PIT count. Also, street outreach workers may be more likely to engage with some individuals if programs target certain populations or geographic areas. Such differences in inclusion probabilities bias the multiple-list estimates downward. [10] This problem, nevertheless, can be mitigated in two ways. The first option would be to modify the procedures by which PIT counters and street outreach engage with the homeless in ways that seek to connect to all populations equally. To the degree those efforts are unsuccessful, there are statistical models or stratified sampling methods that can also address the bias.

Dynamic Estimates of the Homeless Population.

These methods can potentially be applied to generate more frequent estimates (e.g. quarterly or semi-annually) of the homeless population based on multiple lists extracted solely from HMIS data. One list, for example, could be constructed with the names of all people who exit HMIS projects to places not fit for human habitation over a 3-month period. The second list could then

[8] Street outreach programs are not universally implemented by CoCs. For that reason, HUD avoids reporting HMIS statistics based on street outreach programs to avoid skewing comparisons between CoCs. For example, a CoC with a robust street outreach program would appear to have a larger homeless population simply because more individuals are enrolled in the HMIS.

be constructed to include all literally homeless people who enroll in HMIS projects in the next 3-month period. Each list constitutes a sample from the homeless population, and their overlap can be used to generate a population estimate. The one complication is that the assumptions of the two-list method are violated by this approach. The population is not likely to be stable over the full six months of data collection, and the selection probabilities for the two lists are not independent. Nevertheless, statistical models can estimate and control for these interdependencies. These methods, however, require additional research to assess the most appropriate models and to calibrate estimates against other population estimates.

Implementation

An advantage of the multiple-list method is the simplicity and low cost of its implementation. CoCs that already support street outreach programs have access to the necessary data, and the statistical calculations to produce the estimate entail minimal effort. For CoCs that do not have street outreach programs, this method provides an additional impetus to create such programs. Researchers have found that street outreach is an effective and cost-efficiency method for moving unsheltered people into housing. [16] Gaining access to higher quality data on the size of their homeless population would then be an added benefit.

To match the two lists, CoCs need to capture identifying information for each person interviewed in the PIT demographic survey. Such information is often already included in demographic surveys but would be needed to be added if it is missing. Multiple-list estimates only use information on individuals included in the demographic survey and only need to collect a sample of individuals experiencing homelessness. Consequently, CoCs that employ the multiple-list method have the option of redirecting PIT resources away from efforts to enumerate all people experiencing homelessness and focus more on capturing surveys from a randomly selected group of people.

A CoC should also ask each of its street outreach workers to verify that all of their current clients are living unsheltered on the streets the night of the PIT count. It is possible for street outreach clients to enter into a shelter or resolve their homelessness prior to exiting from the program, and these individuals should not be included in the two list estimates.

To minimize bias the CoC should ensure that both street outreach practices and the street demographic survey process sample broadly to include all potential populations of people experiencing homelessness.

Once the two lists are gathered, the CoC needs only to cross-reference the lists to identify those individuals that appear on both. The estimate of the total population is then easily estimated with the following formula:

$$T = \frac{N_s \cdot N_p}{K}$$

Where: T is the estimated total unsheltered homeless population
N_s is the number of individuals on the street outreach list
N_p is the number of individuals on the PIT count list
K is the number of individuals that appear on both lists

The multiple-list estimate can then be employed to check the counts from the traditional PIT survey method. A 95% confidence interval for the multiple-list estimate can be constructed (see the formula in Table 1 of Appendix A), and if the traditional count is performing well, it should fall within the bounds of this 95% confidence interval.

Recommendations

Several recommendations result from this analysis:

- HUD should endorse the two-list method as a valid method for conducting a PIT count and encourage its use.
- CoCs that currently conduct high-quality PIT counts that attempt to cover their entire geography should employ the two-list method to validate the accuracy of their counts.
- Other CoCs that have not been employing methods that strive to cover their entire geography should adopt the two-list method because it improves accuracy even if it is implemented with limited resources.
- CoCs that do not currently support a street outreach program should consider starting one.
- Additional research should be conducted to validate and calibrate multiple-list estimates that rely solely on HMIS data and can be estimated multiple times a year.

Appendix A: Bias and Variance of Estimates of Unsheltered Homeless

There are two goals for estimating the size of unknown populations. The first is accuracy or absence of bias. An unbiased estimate, if conducted multiple times, should on average equal the true number of people experiencing unsheltered homelessness. The second is to minimize variance, the degree to which an estimate varies if the survey was repeated multiple times. Estimates should be unbiased and have as little variance as possible.

HUD in practice accepts PIT counts that employ four different methods.[9] In this brief, we propose a fifth method, the multiple-list method. CoCs may choose to combine some of these methods:

1. **A full census** in which counters visit every region of a CoCs geography on the night of the PIT count and strive to count all people who are living unsheltered on the streets.
2. **Geographic sampling and extrapolation** in which a CoC sends counters to a random sample of geographic areas. These counts are then extrapolated to provide an estimate of the population in the entire region. For example, if the CoC has 10 areas but counters are only sent to 5 randomly selected areas, then the count for the sampled areas is multiplied by 2 (e.g. 10/5) to estimate the number of homeless in all areas.

   Typically, a CoC would employ a stratified sample that differentiates hot areas with known and large homeless populations, warm areas with known but smaller homeless populations, and cold areas where there is no known homeless population. Sampling focuses on hot and warm areas, likely sending counters to all or most of these areas. In contrast, a smaller proportion of cold areas are sampled.
3. **Known location sampling** where a CoC surveys the community prior to a PIT count to learn of the locations with known homeless populations. The PIT count then sends counters to conduct either a full census of those known locations or to a random sample of those locations. If only a sample is counted, the CoC should then extrapolate to the full set of known locations.
4. **Service-based count** where the CoC interviews clients of social service providers and asks them whether they slept in a place not meant for human habitation on the night of the PIT count.
5. **Multiple-list method** where researchers take two or more random samples from the homeless population and estimate the size of the total population based on the overlap of the two samples.

In practice none of the methods is perfect. Resource constraints and the inherent difficulty of counting a population many of whom wish to remain hidden make it extremely difficult for homeless counts to include all unsheltered people. In most cases, known deficiencies in all of these count methods lead to undercounts. Nevertheless, the methods differ in their strengths and

[9] Technically, the known location and service-based counts do not meet the requirements HUD has set out to account for all areas of a CoC. Nevertheless, it accepts these counts recognizing the limited resources many CoCs have to conduct these counts.

weaknesses and the practicality of addressing their limitations. Comparisons of the accuracy and variance of these methods are as follows:

Accuracy. A full census or a sample with extrapolation are the most accurate methods, and if conducted correctly they, in theory, provide unbiased estimates of the total homeless population. Nonetheless, there is evidence that these methods miss many individuals experiencing homelessness. Undercounts arise due to systematic sampling errors where not all locations are canvassed. For example, a homeless person sleeping in a shed on private property will be missed by street counts. Similarly, errors can occur due to measurement error, where people experiencing homelessness are not identified as homeless. Both of these errors tend to bias counts downward.[10]

Multiple-list methods are also unbiased if conducted correctly. Nevertheless, if certain homeless subpopulations are less likely to be included in one of the lists, the estimates are also biased downward. For example, if youth who are experiencing homelessness are more likely to avoid counters, the estimate would be lower than the true population.

Known location and service-based methods are biased downward by design. In each case, the count ignores large portions of a CoC's territory where people experiencing homelessness can reside. Service-based counts are also biased downward because they miss all individuals who are homeless but do not seek out services on the days following the PIT count. Additionally, both known location and service-based counts suffer from the same biases caused by systematic sampling errors and measurement errors that bias census and sampling and extrapolation methods.

Variance. The degree to which these methods produce counts that vary around the true population number differ. A full census, in the absence of systematic sampling or measurement error, has no variance, meaning it should count the same number of homeless in repeated trials. Systematic sampling error and measurement error does lead to variance in estimates. For example, when a census is conducted with more volunteers, it often counts a larger number of people experiencing homelessness because the added volunteers are better able to survey their count areas thoroughly.

Estimates from a sample and extrapolation have large variances. For example, in a 2005 PIT count of Los Angeles researchers divided the 2054 census tracts in the County into hot and cold areas.[17] 244 hot census tracts were sampled with certainty, and among the rest, approximately 15% of the tracts (244 out of 1810) were included in the count. This sample produced an estimate with a 95% confidence interval that ranged 15% above and below the estimate. The estimate was 64,386 and the confidence interval was +/- 9652.

The margin of error produced by a sample and extrapolation can be reduced by increasing the number of census tracts sampled. Nevertheless, there is a tradeoff between larger sample sizes

---

[10] In practice it is also possible for errors to overestimate the number of homeless. Individuals experiencing homelessness may be counted more than once in the count or people who do have homes but who are out on the streets may be mistaken for being homeless. In practice, however, these errors appear to occur much less frequently than the errors that lead to under counts.

and measurement error. If the number of volunteer counters remains constant, increasing the number of census tracts in a PIT count requires each volunteer to cover more territory which most likely increases the number of people experiencing homelessness who are missed.

A multiple-list estimate also has a large variance, but under reasonable assumptions, the variance is likely to be smaller than that produced by a sample and extrapolation given the same number of volunteers. Table 1 shows two hypothetical estimates for a community in which 1000 people reside unsheltered. With 240 volunteers this community could send out 60 4-person count teams. Table 1a provides an example of the results of a sample and extrapolation count where an equal number of teams are sent to hot, warm, and cold areas. The number of homeless identified in the hot zones is 400, in the warm it is 200, and in the cold it is 40. After extrapolations are made, the estimated count is 1000 and the variance is 5719, which yields a 95% confidence interval of 852 to 1148.

Table 1b looks at the results from a multiple-list count. We assume that this community has an ongoing street outreach program with 200 clients enrolled. The PIT night count is conducted

Table 1: Comparison of Confidence Intervals
Sample and Extrapolation versus Multiple-List Method

Table 1a: Sampling Census Tracts with Extrapolation (160 Census Tracts in Area)

Strata	# of Tracts Counted	Total Tracts in Strata	Count in Strata	Strata Estimate with Extrapolation	Variance of Estimate[1]
Hot	20	20	400	400	0
Warm	20	40	200	400	4103
Cold	20	100	40	200	1616
Total				1000	5719

95% Confidence Interval: 852 -- 1148

Table 1b: Capture-Recapture Estimate

Count from Street Outreach	200
Count from PIT Survey	300
Overlap Between Two Lists	60
Population Estimate[2]	1000

95% Confidence Interval[3]: 888 -- 1134

[1] The Variance for each strata is calculated employing the formula $var(\hat{\tau}) = (\frac{N}{n})^2 \times n \times var(c) \times \frac{(N-n)}{(N-1)}$ where $\hat{\tau}$ is the variance of estimated count for the strata, N is the total number of census tracts in the strata, n is the number of census tracts included in the sample, and *var(c)* is the sample variance from the strata. [17] The sample variance assumes that the strata have a bimodal distribution with half of the tracts having no people suffering homelessness and the other half having twice the average number of homeless in that strata.

[2] The estimate of the total is equal to $\frac{N_{SO} \times N_{PIT}}{O}$, where N_{SO} is the number of people enrolled in street outreach, N_{PIT} is the number of people in the PIT count, O is the number of people included on both lists.

[3] The confidence interval for this multiple-list estimate is calculated employing a transformed logit method. [18] It is $N_{SO} + N_{PIT} - O - .5 + \frac{(N_{SO} - O + .5}{O + .5} \times \exp(\pm z_{\frac{\alpha}{2}} \cdot \hat{\sigma}^{.5})$ where the estimated sample variance is

$$\hat{\sigma} = \sqrt{\frac{1}{o+.5} + \frac{1}{N_{PIT}-O+.5} + \frac{1}{N_{SO}-O+.5} + \frac{O+.5}{(N_{SO}-O+.5)(N_{PIT}-O+.5)}}$$

slightly differently because the multiple-list estimate requires that each person be interviewed to collect information to cross-reference person on each list. Consequently, we assume that far fewer people are contacted in the PIT night count, only 300 in comparison to the 640 people counted in the sample and extrapolate method. Given these numbers, if the overlap between the two lists is 60 individuals the estimate of the homeless population is 1000 and the 95% confidence interval is between 888 and 1134, an interval that is 17% narrower than the confidence interval generated by the sample and extrapolation method.

The variance of known location or service-based counts are not well defined because these methods do not rely on random sampling.

References

1. Rossi, P.H., *Down and out in America : the origins of homelessness*. 1989, Chicago: University of Chicago Press. xi, 247 p.
2. Fleming, D. and P. Burns, *Who Counts? Assessing Accuracy of the Homeless Count*. 2017, Economic Roundtable: Los Angeles CA.
3. Hopper, K., et al., *Estimating numbers of unsheltered homeless people through plant-capture and postcount survey methods.* American Journal of Public Health, 2008. **98**(8): p. 1438-1442.
4. National Law Center on Homelessness and Poverty, *Don't Count on It: How the HUD Point-in-Time Count Underestimates the Homelessness Crisis in America*. 2017: Washington D.C.
5. Taeuber, C.M. and P.M. Siegel, *Counting the Nation's Homeless Population in the 1990 Census.* 1990.
6. Clift, T., *What could help Sacramento reduce homelessness? Here's what's being done elsewhere*, in *Sacramento Bee*. 2019: Sacramento.
7. HUD, *Notice CPD-18-08: 2019 HIC and PIT Data Collection for CoC and ESG Programs*, HUD, Editor. 2018: Washington, D.C.
8. Agans, R.P., et al., *Enumerating the hidden homeless: Strategies to estimate the homeless gone missing from a point-in-time count.* Journal of Official Statistics, 2014. **30**(2): p. 215-229.
9. Baiocchi, A., et al., *Homelessness in Sacramento County: Results from the 2019 Point-in-Time Count. Sacramento*. 2019, Institute for Social Research and Sacramento Steps Forward: California.
10. Otis, D.L., et al., *Statistical inference from capture data on closed animal populations.* Wildlife monographs, 1978(62): p. 3-135.
11. Archibald, C.P., et al., *Estimating the size of hard-to-reach populations: a novel method using HIV testing data compared to other methods.* Aids, 2001. **15**: p. S41-S48.
12. Berry, B., *A repeated observation approach for estimating the street homeless population.* Evaluation review, 2007. **31**(2): p. 166-199.
13. Cowan, C.D., W.R. Breakey, and P.J. Fischer. *The methodology of counting the homeless*.
14. Pledger, S., *Unified maximum likelihood estimates for closed capture–recapture models using mixtures.* Biometrics, 2000. **56**(2): p. 434-442.
15. Cormack, R.M., *Log-Linear Models for Capture-Recapture.* Biometrics, 1989. **45**(2): p. 395-413.
16. Mackie, P., S. Johnsen, and J. Wood, *Ending rough sleeping: what works.* An international evidence review. London: Crisis, 2017.
17. Berk, R., B. Kriegler, and D. Ylvisaker, *Counting the homeless in Los Angeles county*, in *Probability and statistics: Essays in honor of David A. Freedman*. 2008, Institute of Mathematical Statistics. p. 127-141.
18. Sadinle, M., *Transformed Logit Confidence Intervals for Small Populations in Single Capture–Recapture Estimation.* Communications in Statistics-Simulation and Computation, 2009. **38**(9): p. 1909-1924.