Session 3A:

Alternative Modes of Estimating the Homeless Population

Estimating Census and Survey Undercounts Through Multiple Service Contacts

Charles D. Cowan
Opinion Research Corporation

Abstract

During the past decade a tremendous amount of work has been done on studying the homeless. The 1980s saw many disconnected city-specific counts of the numbers of homeless that sought to obtain some information on their characteristics. As these studies began to be published, there was a growing awareness of the inadequacy of the methods available and the variety of interests that prompted the studies. The 1980s also saw the growth of new methods and models for counting and the coordination of efforts for multidisciplinary studies. This paper describes the development of new models for counting and new methods of sampling in time and space.

Introduction

The primary focus of this paper is evaluation and estimation. Much of the research that is being conducted on counting the homeless is necessarily concerned with how counts are made. Many different methods have been used in different cities to count the homeless, with varying degrees of success. The ways these methods are developed reflect constraints of the area being counted, the cost of the operation, problems with definitions, and the use to be made of the data by the investigators.

These methods, however, traditionally do not allow for evaluation of the accuracy or completeness of the counts. Furthermore, several of the methods used in the past rely heavily on assumptions that may not be valid and could be tested, or at least studied, in terms of the sensitivity of the model. Finally, most studies have relied on the conduct of a self-contained survey that uses only the data collected as part of that effort—there is no appeal to administrative data available from social services, even though these data could be quite useful.

Methods for data collection have improved. It is possible to enhance the estimation process with supplemental data to improve the accuracy and reliability of the data in an extremely cost-efficient way

and, at the same time, provide an evaluation of the quality and completeness of data from any one source. If evaluation were the only goal of these endeavors, few studies would ever adopt their techniques because of concerns for cost and the usefulness of the data. But the methods to be discussed here have three goals: improvement of the basic homeless counts, reduction of the costs of collection, and provision of information to evaluate the results. In addition, this paper will address some improvements in sampling techniques introduced specifically for these types of problems.

Current practice

Several sources of contacts with the homeless could be used as part of the counting process. Most surveys of the homeless attempt to do a direct count. These surveys or censuses use a mixture of methods, but are usually trying to get a "snapshot" of the population. A snapshot here means a count at a single point in time. These counts are usually done as a census of missions and shelters within an area accompanied by a street count in a sample of blocks in the city. Some studies conduct the census of missions and shelters, and then adjust this count with the ratio of street-to-shelter homeless, using a ratio obtained from some other source.[1] An example of a census accompanied by a survey of street homeless is Frankel.[2]

A carefully designed study involving a census plus a street survey can be very expensive and is subject to several problems that work against an accurate count. Two of the methodological problems that most plague this type of count are the definitional problems of determining who is homeless and the problem caused by the transiency of the population (i.e., while the count is being conducted, the homeless population can be moving, leading to some double counting).

Almost all the studies conducted in the past decade have taken for granted that an individual found in a shelter or mission is homeless. The definitional problem comes from counting individuals on the street. Most counts are conducted at night to minimize the movement of the homeless population during the count and to facilitate screening. However, information for screening and determining who is homeless is not always collected from individuals. It is also true that many homeless sleep in hiding places that do not qualify as housing of any sort, but are places researchers cannot find or reach. The number of individuals who cannot be found does not have to be very large to introduce a significant bias into some estimates.[3]

For street and shelter counts that are conducted in the day or evening, the second problem is that the homeless population is highly transient. Shelter counts can be inflated slightly even during an evening, because the homeless may come in only for a meal and not to sleep, preferring to go back on the street or to a different shelter. Those found on the street, especially in a clustered sample, may be moving between sample blocks during the time the study is being conducted.

Studies that consist solely of a one-time shelter census that is then adjusted for the street population have an additional problem, namely that the ratio being used may be wildly off for the area being counted. The ratio of street-to-shelter homeless can be as low as 1 to 2 or as large as 4 to 1, depending on the area. Without some basic research in the city where the shelter census is being conducted, it is impossible to estimate this ratio accurately.

Any approach to counting the homeless will have methodological problems.[4] The issues described above are presented not as an indictment of the current procedures used to count the homeless, but as a basis for understanding the use of modeling as part of the counting process. An additional reason for considering modeling as opposed to a straight census/street survey is cost. The harder one works within the framework of the census/street survey to correct for methodological problems, the more expensive the study will be. Furthermore, the additional expenses grow much more rapidly than the gains in accuracy. Anything that can be done in a cost-effective fashion can be a major contribution to the count.

A more severe problem with the census/street survey approach is the changing nature of the population. The homeless population is dynamic, and snapshot surveys give a limited understanding of the true nature of the population and its characteristics or needs. A snapshot survey gives no indication of the changes in the size of the population, of whether or not the population observed is itself stable, or of the true demand for a variety of services. Because most of the federal and state studies of the homeless are motivated by the need to know the potential demand for services, it seems strange to use a counting technique that ignores the highly variable nature of this demand.

Use of modeling techniques for counting the homeless

Because of these concerns about the changing nature of the homeless population, it seems natural to turn to the use of modeling techniques that actually capitalize on the transitory nature of the homeless population. These techniques permit both a snapshot view of the population and a measure of the flow of people into and out of the homeless population. The snapshots can be obtained at multiple times in the counting process and can be used as another indication of the changing nature of the population. Finally, modeling techniques can be used to partition demand into measures of constant demand for services, rather than peak or variable demands. The modeling techniques can also be used to determine where there are overlaps in the demand for services across different service agencies—the "service packaging" issue. Finally, a well-designed model will provide information for the evaluation of the estimates generated.

The next sections describe two techniques that use service provider information. One technique is the use of capture-recapture models; the second is the use of surveys that take advantage of the flow of the population. Both techniques rely on models for estimation of the size of the population, and both give considerably more information on the characteristics of the population than the standard census/street survey. Furthermore, both techniques can be used in conjunction with the census/street survey to provide better estimates and an evaluation of the coverage of the census approach.

In particular, model-based approaches can give researchers an estimate of the length and frequency of homelessness episodes; an estimate of flow parameters (such as transition rates for the event of becoming homeless or leaving that state); differences in these characteristics by race, age, and other variables collected as part of the study; and rates of use of services, especially rates of packaging of services. The latter statistics are usually computed by dividing the number receiving services by the estimated number of homeless in the same city, but these statistics often come from two or more disparate sources. The number of service users as a proportion of the total homeless population would be a more precise rate, simply because the estimates would be based on the same samples (allowing researchers to capitalize on the covariance between the numerator and denominator).

Both techniques have been attempted in the past. Some research using capture-recapture methods for homeless populations has been conducted in Baltimore and in Sydney, Australia. The survey

approach has been used primarily with migrant populations (such as farm workers), but would be readily adaptable to the homeless population.

Capture-recapture methods for the homeless

Capture-recapture methods for counting the homeless are, in essence, attempts to use a model that describes how (or with what probability) individuals are observed. Capture-recapture methods have been in use since the 16th century, but have been used with human populations only recently. Capture-recapture might be useful with the homeless population, because the model assumes that in two or more capture periods, every individual in the population has some chance of being observed. By building on the frequencies and patterns of observation for individuals, researchers can estimate the total number of individuals who are actually in the population. However, several assumptions are required for the model to predict the total population accurately, and these assumptions, if violated, can wreak havoc with the estimation process. The description of capture-recapture that follows also provides some discussion about the assumptions required and how they might best be met.

Many shelters and missions keep records of who has been admitted for an evening. In Baltimore, where a study was conducted using capture-recapture techniques, missions and shelters kept complete records, because they were needed for administrative and funding recipiency purposes. The records were not lengthy, but did contain some useful information, including name, age or date of birth, sex, race, and social security number. These data were usually recorded every time an individual stayed in a shelter, so they could be used to track individuals' use of shelters over time. If researchers follow the convention that anyone who stays in a mission or shelter is homeless, at least for that evening, then they can acquire a partial count of the homeless in a city or county by simply collecting admission records from all the missions or shelters. This count employs exactly the same methodology as the shelter count reported in the section entitled Current Practice.

The problem with such a partial count is that it tells nothing about the actual practices of the population in terms of consistent or repeated use of services. These services would include not only the missions or shelters but also other social services, such as outpatient care, soup kitchens, support from the judicial system, job placement, and so forth. Furthermore, it is likely that researchers

using such a count do not even know what proportion of the overall count is represented by the shelter count, and would have to guess, borrow from another study of a similar area (or of the same area in the past), or conduct a street survey to get an accurate measure of this statistic.

Generally, however, it is possible to model the total number of homeless in an area with the amount of information available from the record systems of the missions and shelters, if a few assumptions are made about the population. The missions and shelters have collected the same information on individuals over time. In Baltimore, this information was fairly consistent in most cases, because the shelters needed the information to report to federal and state authorities. Even partial information, if collected consistently, can be useful in studying individuals over time. By using the records of shelters, researchers can match over time to determine which individuals have used the shelter once, twice, three times, and so on. Matching over time in this case means that records are compared for all missions and shelters in an area from time 1 to time 2; thus, a match would exist over time even if an individual stayed in a different shelter each time.

Two important issues arise here. The first is that Baltimore may be unusual in that most of its missions and shelters had fairly complete systems for keeping track of who stayed in a mission or shelter overnight. The systems were not extensive, consisting in most cases of a box of index cards. The information was consistent, however, and that contributed to its usefulness. The homeless routinely gave this information whenever they stayed in a mission or shelter—it was not unusual to be asked for the information, and it was collected in a nonthreatening way. As a result, the data on the cards were complete, with about a 15 percent item nonresponse overall. Social security numbers were consistent within each shelter and across shelters. Spellings of first and last names were less consistent, but this was likely a function of the shelter operators' trying to phonetically interpret what they were told, rather than having the individuals write down their own names.

The second issue is the actual availability of such records. If a particular mission or shelter in a city does not keep such records, it should be possible to ask the mission or shelter operators to keep records for just the nights when the count is being conducted, perhaps offering them an incentive. Again., the records needed are not extensive and the collection should be fairly unobtrusive. However, it should be impressed on the operator that both the accuracy and the completeness of the data are important. In the Baltimore study,

the quality and completeness of the data were fairly impressive, which was a pleasant surprise to the researchers; they had anticipated having to do much more work with the missions and shelters to get quality data.

When data are incomplete or inaccurate, then problems arise in matching records over time, which is the key to the analysis. A brief discussion of the analysis follows, in which the importance of the ability to match over time should become clear.

For the analysis, when matching is complete, a count can be constructed of the number of times an individual has been observed and when (see table 1). These counts can then be used in a model to obtain an estimate of the total size of the population (N).

The value N is the number of homeless who lived in the city during the time the study is conducted. The researchers are assuming that there was a closed, constant population (i.e., no persons entered or left the population during the study period). In table 1, matched records between time period 1 and time period 2 are used to determine the number of homeless seen on both occasions. The total number of persons observed at time period 1 and at time period 2 is also known from the records. If the number of homeless who resided in a city were constant (which would be approximately true for short time periods, such as a few days) then those people who

Table 1. **Observations of Persons at Shelters in Two Time Periods**

	Time period 2		
Time period 1	Stayed in shelter	Did not stay in shelter	Total
Stayed in shelter	M	F-M	F
Did not stay in shelter	S-M	U	U
Total	S	U	N

Note: Values within the cells are counts of the number of times an individual has been observed both times (M), the first time (F), or the second time (S). U = unknown.

were not in the shelters during the first, second, or both time periods would of necessity be those who were on the street. The capture-recapture model would allow for deriving an estimator of N, using just these counts and two assumptions: that the observations between time period 1 and time period 2 are independent and that all homeless persons in a particular capture period have the same chance of being observed in a shelter. This estimator is quite simple and takes the form

$$\hat{N} = \frac{F \times S}{M} .$$

This estimator is not particularly interesting, however, because it makes little use of the records for a larger number of days, the assumptions of independence between the two capture periods and the homogeneity of the observation probabilities are not likely to hold, and there are other sources of information that could be tapped to make the model richer.

There are several ways of expanding this model. First, more days of observation can be used, so that the table is a multiway table, with one dimension for each of the days for which records are acquired. If this is done, the assumption of independence between the observations at different points in time can be relaxed. Higher order models (three-way and higher) allow for assuming a correlation between the observations for any combination of observation days. In the case of the homeless, one would expect to see many of the same individuals repeatedly, because some missions and shelters have the policy of allowing individuals to stay for set periods of time (such as a week or a month) rather than registering anew each day.

The use of higher order models gains something in allowing for relaxation of the assumptions underlying the model being used, but something also is lost, because another of the assumptions being made, namely the assumption that the population is closed (that there are no new homeless and no homeless who leave the population or the city), is less likely to be true the more time periods are included. If data were to be used from time periods spread out over a year, this assumption would no longer be tenable.

The other key assumption mentioned, namely that the probabilities of observing an individual are the same for all individuals at any one capture time, can also be relaxed by stratification of the estimates. This stratification would require tabulating each of the

subgroups in the population (e.g., blacks, whites, and others) separately and making estimates for these subgroups separately. Stratification of the estimates will improve the accuracy of the overall estimate of the number of the homeless within a city, and also will yield some useful and interesting information on subgroups, regarding their propensity to be observed in a shelter and the amount of time they are found in the homeless population.

Stratification will partially resolve some of the problems of heterogeneity of the probabilities of observing an individual. Two problems still exist with the use of capture-recapture estimates: first, there may be some individuals who would fit the definition of homeless who would never use missions or shelters, but would use other social services; second, individuals have different ingress and egress rates from the population. There are capture-recapture models that allow for births and deaths in the population, but these are treated as one-time events. In other words, once a birth is observed, that person is a member of the population until he or she dies; then that individual leaves the population forever. In the larger scheme of things, this seems perfectly reasonable. But for measurement of the homeless population, this is a problem, because a large proportion of the population continually moves in and out of the state of being homeless.

Both of these problems can be addressed by expanding the model still further, beyond a time dimension and the use of a set of stratification variables, to the use of other sources of names of homeless people and interviews at other sites. Research by the University of Illinois on counting rare populations has shown that almost all homeless individuals use some form of social services, even if they do not use a mission or shelter.[5] By intercepting homeless individuals at these service sites (e.g., soup kitchens or medical facilities), one finds that almost all homeless people have a chance of being counted, and in fact have some chance of being counted multiple times within one or several time periods. The same expanded capture-recapture models can be brought to bear on this broader base of information, as long as similar identifying information can be obtained from interviews with these individuals.

At these sites, it will be necessary to conduct interviews for two reasons, Experience in other studies has shown much less consistency in the types of information collected at different types of sites than is found in information collected at missions and shelters. To fit in with the requirements of conducting a matching study, a brief interview or data collection effort with the service provider might be necessary. More important is the need to screen individuals at soup

kitchens or through service providers to ascertain that they meet the definition of homeless. It may seem that the requirement for interviewing weakens the case for capture-recapture studies because it would increase the cost of the survey, but there are actually two cost savings incurred in this type of research. It is true that not all of the information can be collected indirectly, as it can be in many missions and shelters, but some information will be available from the service provider. In addition, even if none of the information is immediately available from service providers, the data can all be collected in one place (the soup kitchen or the site of the service provider), during the day, in a relatively safe environment. The street survey, as commonly conducted, can require more effort and travel on the part of an interviewer and occurs in a less controlled environment.

In interviews conducted at these sites (and interviews with a subsample of individuals at the missions and shelters) it would also be possible to ask questions about where the individual was living the week or month before. This information could be used to estimate parameters in a flow model so that it would be possible to determine when individuals, for the entire population, are in or out of the homeless population. These interviews are necessary, in any event, because it is no longer possible to assume that all individuals seen at social services meet the definition of homeless used for the study. Many people using soup kitchens or other types of social services would not be considered homeless; thus some screening of these populations is also required.

This expanded model can be used to produce snapshot estimates, but it can also produce estimates of the total number of homeless annually in an area such as a city. This estimate would include the chronically homeless, those who are intermittently homeless, and the homeless who are simply passing through a city and thus would be counted once only, or those who temporarily fall on hard times and then recover. None of the other procedures currently in use can produce annual estimates or differentiate between these types of populations.

Evaluation using capture-recapture models

If one of the sources in the capture-recapture model is the combined census/street survey at a single point in time, the census/street survey can be evaluated as a subsystem in the overall model. The overall capture-recapture estimate will give a much more descriptive picture of the population, and homeless who were counted in the

census/street survey could be matched to other captures at other times and with other sources, such as the soup kitchens.

If the census/street survey estimate were C, and the capture-recapture estimate were N, then the ratio of C to N would give an indication of the undercoverage of the census/street survey if all the assumptions stated above hold. The census part of the census/street survey would be precisely the same as the first count obtained for the capture-recapture model. The additional count from the street survey would be the portion estimated by conducting a second count at a later time. The street survey would be represented in part by those counted at time 2 but not at time 1, and in part by the estimated portion of the homeless who were not counted at time 1 or time 2. If the assumptions hold, this should also be an unbiased estimate of the number of persons who were homeless at the time of the study. The capture-recapture is not subject to sampling error, but does have a modeling error component. If the street survey is subject to some underreporting because persons on the street are not identified correctly as being homeless, or are not found because of an inadequate search, then they are undercounted in the census/street survey, but are represented in the capture-recapture estimate because of their chance of inclusion in that estimate.

It may be especially useful to further analyze the data collected for the capture-recapture model to determine whether those captured in the census/street survey were consistently found primarily in missions, through other social services, or both. This analysis will indicate the efficacy of the street sample, which is the most costly and difficult part of the operation to perform.

Following such an analysis, it may also be useful to determine the cost of collection using different sources. This type of analysis would help researchers in the future determine where their research monies are best spent and what sources of information would have the best payoff. Such an analysis would also indicate whether researchers should even consider taking street samples, because the overall utility of taking snapshots of this sort is low in relation to the volumes of data available from a modeling approach.

Sampling in time and space

The final section of this paper will introduce an alternative methodology for counting the homeless. These methods are grounded much more in sample survey theory than in modeling. The primary impetus for these methods has come from studies of migrant workers

and nomads, but the methods are adaptable to studies of the homeless and are potentially more realistic than capture-recapture methods in terms of their ability to find and count the homeless.

The sampling methods developed for these "nonsedentary" populations involve stratifying the sample to permit sampling in both time and space. Censuses and street surveys attempt to deal with the problem of the mobility of the population by doing the entire count of the population in one evening, preferably overnight, to minimize double counting and maximize the researcher's ability to find individuals. When sampling in time and space, stratification of the time dimension makes it very difficult to double-count individuals if the time strata are short. Furthermore, by stratifying and establishing appropriate counting rules, the objective becomes one of counting individuals in a particular place at a particular time. Counting individuals in more than one place at different times is accepted under this sampling model and can be dealt with in the estimation procedure, so that an unbiased estimate of the number of individuals in the population can be obtained.

If relatively small areas are established for sampling, then individuals who enter these areas (for example a particular block or a line at a soup kitchen) can be sampled during the time interval established for that stratum. Over time, the sampling areas can change. In some versions of this method, a time-space grid is established and cells in the grid are sampled, so that a number of specific locations and specific times are sampled with replacement. In other versions of this method, the locations are divided into strata and sampled; then, within each of the sampled locations, the time intervals are sampled as clusters, yielding a two-stage sample. There are other variations on these sampling schemes. The point is that the use of methods for sampling in time and space reduces the need to conduct the street survey in a single night, making the operational aspects of conducting the count easier, though the estimation procedure becomes a bit more complicated.

One additional set of strata for the survey needs to be established to make it possible to cover the entire population. This would be the set of strata including members of the population who are not found in any of the physical strata being used for the survey. These strata may be used for individuals who are in transit between sites during one of the sampled times, for example. Estimation proceeds in the usual way for a sample survey, except that it is performed across both demographic/geographic strata and the time strata. Under some sampling models, it may be preferable to use multiplicity or

network sampling methods to correct for overcounting of some individuals in making an estimate of the population as a whole.

Kalsbeek has developed several estimators in which sampling is done in time and space to measure the size of migrant populations, such as farm workers.[6] In a 1986 paper, Kalsbeek also considered sampling designs of this type for use in the study of nomadic populations in Somalia.[7] Kalton has looked at sampling flows of mobile human populations more generally, and has developed several designs for estimating the number and characteristics of users of museums, concert attendees, and similar highly mobile groups.[8] His methods parallel those of Kalsbeek and would be adaptable to the homeless.

Although not as rich as the modeling approaches described in the section on capture-recapture methods, these sampling methods would be extremely useful in providing a solid methodology for counting the homeless population and determining its characteristics. One major advantage of these sampling approaches over the capture-recapture methods is that fewer assumptions need to be made for the modeling process. The sampling rules are straightforward and easy to use.

Finally, with this approach, the census/street sample can again be a subset of the overall sampling and survey effort. Estimates from the census/street sample can be compared with additional data collected by sampling in time and space to determine the extent of undercoverage of the census/street sample. Because the approach involving sampling in time and space also allows interviewing in places such as social service administration centers, it permits broader interviewing and a better determination of who the homeless are and the types of places these individuals may be found.

Conclusion

Two basic approaches to counting the homeless have been proposed, both involving broader and more wide-ranging sampling practices than those currently used. The potential advantages of either approach over more traditional methods are legion–the information obtained from the models, the improvements in coverage, and the ability to evaluate the quality of the work done make both procedures superior to the simpler methods of conducting a shelter census and/or a street census. Both approaches, if conducted properly, have the potential of being less expensive. In the 1990s, more

research will be done using system approaches (combining sampling with administrative record systems) and the use of models for research on the homeless.

Author

Charles D. Cowan, a mathematical statistician, is the chief statistician and vice president for advanced analytics for Opinion Research Corporation, Washington, DC.

Endnotes

1. A discussion of the ranges of the street-to-shelter ratios and their use in estimation can be found in Charles D. Cowan, William R. Breakey, and P.J. Fisher, "The Methodology of Counting the Homeless," in *Homelessness, Health, and Human Needs* (Washington, DC: The National Academy Press, 1988).

2. Martin R. Frankel, "A Probability Sample of the Homeless Population of Chicago," in *Proceedings of the American Statistical Association,* Survey Research Methods Section (Alexandria, VA: American Statistical Association, 1986), 176-77.

3. Charles D. Cowan et al., *Evaluating Censuses of Population and Housing* (Washington, DC: Bureau of the Census, 1985).

4. Charles D. Cowan, William R. Breakey, and P.J. Fisher, "The Methodology of Counting the Homeless," in *Proceedings of the American Statistical Association,* Survey Research Methods Section (Alexandria, VA: American Statistical Association, 1986), 170-75.

5. A discussion on service use can also be found in Cowan, Breakey, and Fisher, "The Methodology of Counting the Homeless."

6. William D. Kalsbeek, "Sampling in Time and Space for Surveys of Nonsedentary Populations." Paper submitted to *Survey Research Methodology,* Ottawa, Canada: Statistics Canada (1991).

7. William D. Kalsbeek, "Nomad Sampling: An Analytic Study of Alternative Design Strategies," in *Proceedings of the American Statistical Association,* Survey Research Methods Section (Alexandria, VA: American Statistical Association, 1986), 164-9.

8. Graham Kalton, "Sampling Flows of Mobile Human Populations," in *Proceedings of Statistics Canada Symposium 90: Measurement and Improvement of Data Quality* (Ottawa, Canada: Statistics Canada, 1990).