

1. INTRODUCTION

There are very few local studies which estimate the number of homeless for a city or county. There are even fewer which make state or national estimates of the number of homeless. This paper examines the methods used to count the homeless, presents a description of a research study being conducted in the city of Baltimore, and reviews the problems encountered in trying to make an estimate of the number of homeless. The purpose of this paper is not to promulgate a particular method of counting, but rather to explore why counting the homeless is such a difficult and expensive task.

2. REVIEW OF THE LITERATURE

Although there is a great deal written about the homeless (primarily in the form of case studies and review of secondary data sources), there are very few papers or reports that deal with the number of the homeless. This review will only look at the literature that deals with the numbers of homeless, methods for counting the homeless, or papers on estimation procedures for counting difficult to enumerate or populations.

Three methods are used to develop estimates of the number of homeless:

- 1) indirect estimation,
- 2) single contact censuses, and
- 3) capture-recapture studies.

The indirect method involves eliciting information from knowledgeable sources about the number of homeless in an area or receiving services. This information may often be from agencies or groups that deal with the same sets of people, and so usually some allowance has to be made for double (or multiple) counting of individuals. In addition, the researcher has to ensure that each of the respondents in this type of study is using approximately the same definition of who is homeless in their reports, and that the reports of the number of homeless encompass approximately the same time period. The advantage of indirect estimation is that it is the cheapest method for making an estimate of the number of homeless.

There are two reports of studies that seem to use indirect estimation techniques. The first is Homelessness in America (Hombs and Snyder, 1983). The authors report that: "No one can say with certainty how many people in this nation are homeless. ... in 1980, we prepared a report, for a Congressional committee, on the national dimensions of the problem. At that time, we concluded that approximately 1 percent of the population, or 2.2 million people, lacked shelter. We arrived at that conclusion on the basis of information received from more than 100 agencies and organizations in 25 cities and states.

... It is as accurate an estimate as anyone in the country could offer, yet it lacks absolute statistical certainty." This number, despite the flaws inherent in trying to obtain a national estimate from such a small, disparate, and unrepresentative sample, was the only national number available and so attained a certain level of currency. A more recent study, also designed to produce an estimate at the national level, was conducted by the Department of Housing and Urban Development. The Report to the Secretary of Housing and Urban Development on the Homeless and Emergency Shelters (Bobo, 1984) used a similar methodology in that it was a telephone interview study which contacted 500 knowledgeable observers to obtain local estimates of the number of homeless in a sample of 60 urban areas. The HUD study reports a range of 250,000 to 300,000 homeless in the U.S.. The discrepancy between the two estimates is characteristic of estimates of homeless at all geographic levels. This uncertainty as to the magnitude of the population size makes it extremely difficult to even plan (let alone execute) a direct contact study of the homeless, since budgeting for the correct level of resources becomes a guessing game.

The single contact census is a technique commonly used by cities to make estimates of the size of the resident homeless population and to describe it to a certain degree. The advantages of a single contact census are twofold: the researcher has some direct contact, even if only by observation, with each of the individuals counted within a discrete time period and so the possibility of multiple counting individuals or groups is almost negligible, and the researcher can ensure that the individuals contacted fit the study's definition of being homeless. There are also two primary disadvantages of a single count census. The census is a snapshot of the population at a single point in

Charles D. Cowan is Chief Statistician, Center for Statistics, Department of Education. William R. Breakey is Director of the Community Psychiatry Program, Department of Psychiatry and Behavioral Sciences, The Johns Hopkins Medical Institutions. Pamela J. Fischer is Assistant Professor, Department of Psychiatry and Behavioral Sciences, The Johns Hopkins Medical Institutions.

time, but the homeless population is always in a state of flux and transition. The other disadvantage of a single contact census is that it is expensive relative to the indirect estimation techniques.

The most recent single contact census of the homeless in a city was conducted in Washington, D.C. by the Center for Applied Research and Urban Policy at the University of the District of Columbia. (Robinson, 1985) The report presents a direct count and two estimates of the number of homeless in Washington, D.C.. The estimates differ from one another by 48 percent, reflecting the researcher's assumptions about how well the enumerators might have done at finding the "concealed" homeless population. The full range of estimates of the number of homeless in Washington, D.C. is presented (using five different methods for generating the estimates) is 4347 homeless to 7152 homeless. Other recent single count censuses have been conducted in Boston, Phoenix, and Pittsburgh. (HUD Report on Homelessness, Hearings). All three counts were conducted in 1983. The reports on these studies emphasize that the count was not intended to be complete, that each is only a partial count.

Another type of study that expands on the methodology employed in the single contact census uses multiple contacts. This type of study avoids the problem of characterizing the homeless in a snapshot at a single point in time. A recent study of this type was reported for the city of Nashville (Wiegand), where the homeless were enumerated by a census conducted at four separate times during a year. The report from this study was able to compare the demographic distributions observed at different times and draw conclusions about the changing nature of the homeless population in Nashville. The multiple contact type of study gives the researcher more information on the changes in demographics of the homeless population, but does not give information about why there may be changes. Furthermore, the changes observed are only the gross changes, without information about what is happening to individuals in the population (or net changes).

The last type of study to be reviewed examines individuals in the homeless population over time. Capture-recapture techniques also involve multiple observations on the homeless population, but capture-recapture requires matching observations of the homeless enumerated between each of the data collection periods. Capture-recapture assumes that each data collection is imperfect and that there is some probability at each data collection that individuals will be missed, and consequently there is some

(unknown) probability that individuals will be missed both times (or in the case of a multiple capture census, some individuals will be missed by every census effort). There is only one study of the homeless population that make uses of capture-recapture techniques. This was a study of the number of homeless men in Sydney, Australia (Darcy and Jones, 1975). In their study of homeless men, the researchers conducted a one day census at 25 locations including shelters, hospitals, clinics, and a gaol. The census was conducted three times at the 25 locations on June 30, 1971, October 13, 1971, and March 8, 1972. The authors estimate the number of homeless men in Sydney to be 3200. The authors in this study were concerned that the three month lag between censuses was sufficiently long to allow entry and exit from the homeless population (through moving in and out of Sydney, deaths, etc.). They estimated the average birth and death rates for the homeless to be 21 percent and 5 percent respectively, a strong indication that the homeless population was increasing at the time of the study.

The capture-recapture estimator does require a number of different assumptions to be made, including assumptions about the difficulty a researcher will encounter in trying to find the homeless. The next section will describe a capture-recapture study being conducted in the city of Baltimore, and the following section will review the methodology being used and the assumptions made to plan the study.

3. A CAPTURE-RECAPTURE STUDY IN BALTIMORE

There are 15 missions and shelters in Baltimore city. During the period August, 1985 and May, 1986, there will be eight censuses conducted of the 15 missions and shelters. The censuses will be conducted at the beginning and the end of August and November, 1985, and February and May, 1986. The days of the census in each month are chosen so as to be shortly after welfare checks are issued (in the early part of the month), and for the latter half of the month, some temporal distance from when checks are issued. Each of the missions and shelters keeps records of who stays overnight, and was asked to complete cards designed by the authors on each person who was a resident on a prespecified night. The cards ask for name, sex, race, age, date of birth, social security number, aliases, and other descriptive information that would uniquely identify the homeless persons. The census uses only missions and shelters, and not soup kitchens, hospitals, or lockups because of the impossibility of asking screening questions on the study's limited budget. Occupants of

missions or shelters are automatically defined to be homeless for the purposes of this study. Every record in the final data file indicates how many times each homeless person has been enumerated and where, plus it includes all the demographic information collected at each observation. From these records, estimates are made of the number of the homeless.

4. METHODOLOGICAL CONCERNS - WHAT'S NOT IN THE MODEL

The most important difference between the capture-recapture technique and the other two counting techniques described earlier is that the capture-recapture technique is the only technique that involves a statistical model. With a statistical model, the researcher postulates a mechanism that describes how the data to be analyzed is generated. In postulating the mechanism, the researcher makes a number of assumptions about factors that affect, or perhaps more importantly, do not affect the data collection process. The most important assumptions are listed below:

- 1) Clear definitions - homeless persons can be identified;
- 2) Homogeneous observation probabilities - each person has the same chance of being found in a single period;
- 3) Stability - the size and nature of the population does not change during the observation period;
- 4) Stationarity - the population doesn't move in and out of the study area during the observation period;
- 5) Captures independent - for t time periods, the (t) th order interaction term is zero;
- 6) Data correctness - the information collected on the homeless is complete and without error;
- 7) Matching correctness - data records for the same individuals can be linked between observation periods;
- 8) Single observations - individuals are observed at most one time at each data collection;
- 9) Externalities known - factors that affect the data collection are known and can be accounted for;
- 10) Response complete - homeless individuals or shelters provide information requested.

Violations of these assumptions invalidate the model, causing the results to be biased. The researcher can develop a more complex model that allows more flexibility in the data collection process (i.e., a model that allows for all exigencies), but the researcher discovers rapidly that the more complex model requires much more data than the simple model, or that at times the complex model is inestimable. What follows is a description of the types of assumptions or problems that can affect the estimation process in use of cap-

ture-recapture for estimation of the size of the homeless population, and some possible solutions where they are possible.

4.1 Clear Definitions

We have tried to avoid the problem of defining who is homeless by counting only those homeless observed at shelters or missions and making the assumption that everyone who uses such facilities must be homeless. This may be a bad assumption; it may be that there are people who use the facilities who have some form of support and so would not be considered homeless by other researchers. Second, even if everyone who uses a shelter or mission is homeless, not all homeless use shelters or missions, and so our study design does not encompass the complete population.

There is no universally accepted definition of homelessness. Researchers seem unable to agree on what criteria apply to determine who is homeless, or under what situations a person is homeless. The most commonly used criteria consider three different variables:

- 1) Does the individual have familial or community support?
- 2) Does the individual have a source of income?
- 3) Does the individual have a place to stay?

But each of these questions leads to more questions. What constitutes community or family support? If a person can live with a relative and chooses not to, is that person really homeless? How steady does income have to be before the person is no longer considered homeless? What is a minimum standard for housing before the person is considered homeless? It may be that no rules will ever be available since different service organizations have different target populations that they serve, and strict rules on who is homeless for the purposes of research will be at variance with the qualifications required by these organizations.

4.2 Homogeneous Observation Probabilities

The use of capture-recapture techniques makes an assumption that each person has a chance of being observed, and that being observed is a Bernoulli event with a constant probability for all persons. The probability of being observed can differ between data collection periods; in fact, it is expected to vary from one time period to the next. Imagine a totally stable homeless population that did not change overall from time period to time period. Varying rates of shelter usage would indicate that the researcher should expect a different chance of observing individuals on different days.

A much more severe problem related to the homogeneity of observation proba-

bilities are those cases where the chance of observation is zero for a particular time period, or for all time periods. For the study being conducted in Baltimore, with eight data collection periods, the homeless who make even infrequent use of shelters should be represented in the final estimate. Only those who never use shelters will not be represented. Those persons who make no use of shelters or missions essentially comprise a separate stratum for which there is no estimate and no observations. To correct for this omission, the researcher has to rely on external information. In this case, the researcher has to determine from other sources what proportion of the population never use shelters or missions. This proportion can be used in the same way that the ratio of street homeless to shelter homeless is used to adjust the final estimate.

4.3 Stability

In usual capture-recapture studies, it is assumed that the population remains stable. The homeless population is not stable, however, in that members of the population can move in and out of the population without ever changing their locale. Over a ten month period, as in the Baltimore study, one would expect that the size of the homeless population would fluctuate. There is some number of homeless who move in and out of the population due to a variety of reasons. The problem is that a closed population model assumes that anyone not observed was simply missed, not out of the population. A more appropriate model would allow the homeless population to be an open population, i.e. one that changes size rather than remaining static. Even if the homeless population were not especially mobile, the fact that some homeless move in and out of the population makes an estimate of the size of the homeless population much more difficult, but perhaps also more meaningful. It is necessary first to understand that the idea of a single size for the homeless population is a meaningless concept. This is a changing group, much more volatile than the rest of the U.S. population. The purpose of measuring the size of the homeless population is to in part determine the potential demand for services by a particularly disadvantaged and underrepresented minority. But this is only part of the picture needed for the homeless since the data collected will be out of date by the time the data is analyzed.

4.4 Stationarity

The homeless population is also fairly mobile, meaning that during an extended study period the homeless will be moving in and out of the study. The closed population model assumes that a person not observed is simply missed in

the attempt (or in the case of the Baltimore study, did not stay in a shelter but was still in the area). The problems of mobility and mortality in terms of being in the study area and being in the homeless population can be dealt in part with by using an open population model instead of a closed population model.

4.5 Captures Independent

Capture-recapture techniques require that at some level there is no interaction between observations. For a two sample problem this means that the event of being observed in the first sample for an individual should be independent of the event of being observed in the second sample for the same individual. For a three sample problem, the two way interactions can exist (the first observation can be correlated with the second, the first with the third, and the second with the third), but there must be no three way interaction between the events of being observed in all three samples. For higher order models (those involving more than three samples), all interactions can exist except that representing the highest level. With multiple samples available for the Baltimore study, one would expect that correlated observations would be less of a problem. This is not the case, however, because of the rules of residency employed by some of the shelters and missions where the observations are taken. Some missions allow the homeless to stay for a specified period, and so individuals are almost guaranteed of being found repeatedly if multiple observations are taken in a short period of time. However, ultimately after eight data collections spaced out over ten months the correlation should be nonexistent. We would expect that samples collected in the same month (pairs [1,2], [3,4], [5,6], and [7,8]) should exhibit the most correlation, followed by other pairs of samples and then triplets of samples that are temporally close.

4.6 Data Correctness

The homeless population is an especially difficult population from which to gather information because the homeless are especially jealous of their privacy, a sizeable proportion of the population suffers from some form of mental disability, and because the data collection is done by operators of missions and shelters. In the latter case, the operators have the advantage of having gained the trust of the respondents (in some cases only after a long period of building up a relationship with the homeless person), but the operator's primary focus is not on collecting data needed for the survey and so he may not press for information if he perceives gaining it may jeopardize

relationships with the respondents. Data collected for capture-recapture studies needs to be as complete and correct as possible, since the data serves multiple purposes. The data are used first of all for matching purposes, and the estimation procedure can be especially sensitive to matching errors caused by incomplete, erroneous, or conflicting data.

4.7 Matching Correctness

Capture-recapture techniques are very dependent on having the observations correctly matched between time periods, especially for populations where the probability of being observed in any one time period is low or only of moderate size. Even when the data collected is complete and without error, however, one can still have matching errors or problems in determining which cases are matched. The most severe problem in conducting a match of homeless is incomplete data. In some cases, no information that uniquely identifies an individual is collected at the missions or shelters, making it impossible to determine if the same individual is observed in different time periods. In other cases, the data seems to indicate a match to an individual observed at a different time, but the researcher is not certain of the match. This can occur particularly when some of the data seems to be conflicting over two time periods.

4.8 Single Observations

Another assumption that underlies the use of capture-recapture techniques is that each individual is observed no more than once during a data collection period. The effect of unknown multiple observations is very severe in capture-recapture estimation since each duplicate entry is multiplied by the size of the population observed at other data collection periods. In the Baltimore study the authors had not anticipated this as a problem since each mission was filling out cards from its records and it was unlikely a homeless person would be listed twice in the same mission. What we did not anticipate was that a homeless person may be listed at two different locations on the same night. It may be that the individuals who are at multiple locations may go to the first location to get dinner, but a second location because it may be a better place to stay.

4.9 Externalities Known

There are sometimes factors that are external to the research plan that affect the data collection, either adversely or favorably, but more often the former. External factors may be controlled in some instances so that the effect on the research can be minimized. One such factor in the Baltimore study would be the issuance of welfare checks. By

choosing some of the data collection days to be a set number of days before the dates of check issuance and the remainder of the days a set number of days after check issuance, one can attempt to control for the availability of welfare funds in modelling of the size of the homeless population.

There are external factors that cannot be controlled, but at least they can be measured, and so modelled. In the Baltimore study, in addition to the data gathered in the missions on the people who had stayed in the missions on the dates of collection, one can also collect information on the occupancy rate for each mission or shelter, and also on the temperature and amount of precipitation for that evening. These factors can be used in modeling shelter usage and in determining the likelihood of particular groups of individuals using the missions.

4.10 Response Complete

This category is somewhat different than category 4.6 on data correctness, because that category assumed that responses had been obtained from operators and the homeless who registered with the shelters or missions under study. There will be times that the homeless do not provide information upon registering in the shelter or mission, so that there is insufficient information to determine anything more than that there was a person who stayed in the mission. The records for these individuals cannot be matched over different data collection periods, but it is entirely possible that these individuals were at the same or a different mission at other data collection times. At this time it seems that the best way to deal with these cases is to provide a noninterview weight on all cases in the analysis, dropping the cases with no information for matching and weighting up those cases that are able to be matched to represent themselves and the cases that are dropped, perhaps with different weights for each mission.

A more severe problem can be encountered if some of the missions do not agree to provide the information needed for the study. Since individuals can use the missions for extended periods of time, it may be that some individuals will be in the missions not providing information, which means they have no chance of being included in the study. Obtaining a count of individuals in the nonresponding missions would at least provide a lower bound for a single time period of the number of homeless, but for multiple time periods the counts of shelter users would have to be incorporated in the noninterview weighting as described in the previous paragraph. In the Baltimore study, only one of the missions did not cooperate.

5. PLANS FOR DEVELOPING RESEARCH AND CONCLUSIONS

There is much work that needs to be done before the final estimates are available from the Baltimore study. Four specific areas where developmental work is needed can be listed:

- 1) There is a need for a more comprehensive model. Data from several sources are being used and adjustments for missing data need to be incorporated into the model.
- 2) There is a need for more data from other related studies. Research on other homeless studies needs to be conducted and coordinated so that some of the adjustments called for in this paper can be effected.
- 3) There is a need for more comparative research to determine better methodologies for studying difficult to find or difficult to enumerate populations.
- 4) There is a need for more cooperation between agencies and researchers to develop a workable definition of who is considered homeless. Little useful research can be conducted on a national level until agreement is reached on what the population encompasses.

For the study described in this paper, the authors expect that a full analysis of the data collected in the city of Baltimore will take some time. A complete analysis of the data will include determining the effect of incomplete or inconclusive data for matching purposes, and an analysis of the level of correlation among observations between data collection periods.

6. REFERENCES

- Bishop, Yvonne M.M., Fienberg, Stephen E., and Holland, Paul W. Discrete Multivariate Analysis. The MIT Press, Cambridge, MA, 1975.
- City of Chicago, Department of Human Services. Homelessness in Chicago. Chicago, 1983.
- Cowan, Charles D. The Effects of Misclassification on Estimates from Capture-Recapture Studies. Unpublished doctoral dissertation, The George Washington University, 1984.
- Crystal, Stephen and Goldstein, Merv. Correlates of Shelter Utilization: One-Day Study. Human Resources Administration of the City of New York, New York, 1984.
- Darcy, L., and Jones, D. L. "The Size of the homeless men population of Sydney." Australian Journal of Social Issues, 10 (1975), 208-15.
- Fischer, Pamela J. and Breakey, William R. Homelessness and Mental Health: An Overview. Report by the Department of Psychiatry and Behavioral Sciences, Johns Hopkins Medical Institutions, Baltimore, MD, 1985.
- Hombs, Mary Ellen and Snyder, Mitch. Homelessness in America: A Forced March to Nowhere. Community for Creative Nonviolence, Washington, D.C., 1982.
- Levine, Irene Shifren. "Homelessness: Its implications for mental health policies and practice." Paper presented at the meeting of the American Psychological Association, Anaheim, CA, August, 1983.
- McGerigle, Paul and Lauriat, Alison S. More than Shelter: A Community Response to Homelessness. Report by United Community Planning Corporation, Boston, MA, 1983.
- Phoenix South Community Mental Health Center. The Homeless of Phoenix: Who Are They and What Should Be Done? Phoenix, Arizona, 1983.
- Seber, George A. F. The Estimation of Animal Abundance. Griffen, London, 1973.
- U.S. Department of Health and Human Services. Alcohol, Drug Abuse, and Mental Health Problems of the Homeless. Proceedings of a Roundtable held March 31 and April 1, 1983.
- U.S. Department of Housing and Urban Development. A Report to the Secretary on the Homeless and Emergency Shelter. Office of Policy Development and Research, Wash., D.C., May, 1984.
- U.S. General Accounting Office. Homelessness: A Complex Problem and The Federal Response. USGAO, Washington, D.C., 1985.
- United Way of America. Report of the President's Task Force on Food Assistance. Washington, D.C., 1984.
- Wiegand, R. Bruce. "Counting the Homeless", American Demographics, Vol. 7, #12, Dec. 1985, pp.34-37.