Journal of Official Statistics, Vol. 13, No. 1, 1997, pp. 59–73

Issues in the Use of a Plant-Capture Method for Estimating the Size of the Street Dwelling Population

Elizabeth Martin, Eugene Laska, Kim Hopper, Morris Meisner, and Joe Wanderling[1]

In 1990, the U.S. Census Bureau conducted two operations designed to include homeless persons in the census: an enumeration of the occupants of emergency shelters, and a late night enumeration of street sites identified by cities and census offices as places where homeless people congregate. To assess the street enumeration, the U.S. Census Bureau sponsored independent studies in which unobtrusive observers were stationed in a sample of street sites. The observers reported their observations and experiences of the enumeration process in debriefing questionnaires filled out immediately after the conclusion of the street enumeration. Data reported by the observers are applied here to fit a plant-capture model, which is an alternative to the classic capture–recapture method of estimating the size of a population. This method assumes that the plants (in this application, the street observers) have the same capture probability as other members of the target population. The plant-capture method has potential application as a method to evaluate coverage of the homeless population and other populations for which the assumption of closure is questionable. The article analyzes the data to develop various estimates of the capture probabilities, and assesses the strengths and weaknesses of the method as a potential source of coverage estimates in future enumerations of the homeless population. The article also discusses weaknesses and uncertainties in the street observer data, and evaluates how the quality of the data may affect future attempts to base coverage estimates on similar data from observers or plants.

Key words: Unobtrusive observation; dual systems estimator; census coverage; homeless population.

1. Introduction

The immensely complex conceptual and operational problems of counting the homeless population remain unresolved, as does the related question of how to estimate coverage for this elusive segment of the population. However, data collected for an assessment of street enumeration conducted during the 1990 census provide an opportunity to explore models which eventually may yield estimates of coverage of the unsheltered homeless.

On the night of March 20–21, 1990, the U.S. Bureau of the Census conducted a special

[1] This article reports results of research conducted by staff of the U.S. Census Bureau and Nathan S. Kline Institute, funded in part by a Joint Statistical Agreement with the U.S. Census Bureau and by NIMH grant 50193. Elizabeth Martin is affiliated with the U.S. Bureau of the Census. Eugene Laska and Morris Meisner are affiliated with the Nathan S. Kline Institute for Psychiatric Research and New York University Medical Center. Kim Hopper and Joe Wanderling are affiliated with the Nathan S. Kline Institute for Psychiatric Research. An earlier version of this article was presented at the 1993 meeting of the Western North American Region Biometric Society/Institute of Mathematical Statistics in Laramie, Wyoming. The views expressed are the authors' and do not necessarily reflect those of the U.S. Census Bureau, Nathan S. Kline Institute, or New York University Medical Center. We thank Charles Alexander, Larry Ernst, Freddie Navarro, Annetta Clarke, and Laurie Schwede for useful comments.

two-phase ''Shelter and Street-Night'' (S-Night) operation in which enumerators counted people in emergency shelters, street locations, and other places not intended for habitation, for the purpose of including homeless people in the census (see Taeuber and Siegel 1991, for a description of S-Night). During street enumeration, which took place from 2 to 4 a.m., enumerators were to interview all people visible and awake, who were not in uniform or engaged in obvious money-making activities, in pre-identified nighttime street sites and allnight places of commerce. No screening question was asked to determine if a person had a usual home, or was homeless. Sleeping persons were not to be wakened for an interview, but were counted by observation. Each enumeration team was assigned a number of sites to enumerate in the designated 2–4 a.m. period. The sites – city parks, areas under bridges, bus and train stations, and other locations where homeless people were thought to stay at night – had been identified prior to the census by local governmental units, police, groups working with homeless persons, and U.S. Census Bureau district office personnel.

The 1990 census was the first census in which the U.S. Census Bureau attempted systematically to include homeless people on the street in the census, and the S-Night procedures were developed especially for that purpose. Therefore, the U.S. Census Bureau sponsored an assessment to determine how well enumeration procedures were implemented and followed by enumerators at street sites, and to identify external factors that influenced the street enumeration.[2] Researchers at the Nathan Kline Institute were further interested in testing the feasibility of a new statistical approach for adjusting a population count in order to give a better estimate of its true size. This article reports on both objectives.

The method of the assessment, which relied upon reports of unobtrusive observers (who were not enumerators), had never before been used by the U.S. Census Bureau as a method to assess census operations. Researchers in each of four cities placed teams of 60 in-place-observers, and in New York placed 127 observers, at a sample of street sites which had been pre-designated for enumeration. The cities of Chicago, Los Angeles, New Orleans, New York, and Phoenix were chosen purposively to represent different regions and weather conditions, and to include the two cities believed to have the largest homeless populations (New York and Los Angeles). Researchers were given standard instructions for selecting a random sample of all pre-identified commerce and street sites in their study areas, using records prepared by the local U.S. Census Bureau district offices to assign enumerators to sites. In all cities but New Orleans, the study area represents only part of a city, so the results cannot be generalized beyond the specific areas covered.[3]

The method of stationing unobtrusive observers, sometimes referred to as ''plants,'' in street sites originally was proposed (by the second author) as a possible method for adjusting counts to obtain a better estimate of the homeless street population. It was believed that

[2] The S-Night Assessment project was managed by Pamela Campanelli and Matt Salo, who, assisted by Laurel Schwede, planned and implemented the project and designed all procedures within a very short time period. The data were collected and assessment reports prepared under the direction of Kim Hopper (New York), James Wright and Joel Devine (New Orleans), Kathryn Edin (Chicago), Michael Cousineau (Los Angeles), and Louisa Stark (Phoenix).

[3] The study areas covered Manhattan south of 110th St. on the westside and 96th St. on the eastside, part of central Chicago (including the loop), central Los Angeles (including Skid Row area), most of Phoenix excluding the westernmost portion, and Orleans Parish.

the results of the street enumeration could be adversely affected by a number of procedural difficulties caused, for example, by enumerators' difficulties finding or gaining access to sites or difficulties counting the persons found there. If so, then it may be possible to improve the quality of the count by use of an estimation procedure. Although the U.S. Census Bureau sponsored the use of the method in order to gain more knowledge about how well S-Night procedures were implemented, the bureau took the position that it would be inappropriate to draw conclusions about S-Night coverage based on the proportion of observers who were counted, or to use their reports to make adjustments to the counts. Some of the reasons for this position were limitations on the sample of sites and study areas, questions about the possible influence of observers on the outcome of enumeration, and uncertainties about the validity of the observers' reports of whether they were enumerated. Results of the S-Night pilot study permit us to address some of these issues empirically.

2. Methods

2.1. Counting

Based on guidelines provided by the U.S. Census Bureau, observers or plants were trained in census enumeration procedures, and instructed to remain within the site boundaries and to stay in the open throughout the enumeration period so that census enumerators could see and enumerate them. They were neither to approach nor to avoid census enumerators, were to dress and act in a manner that allowed them to blend in with others on the site, and were to remain as unobtrusive as possible. They observed whether enumerators came to the sites and, if so, how they conducted the enumeration. Plants also were to report whether they were interviewed or, if not, whether they believed they were counted by observation. They recorded their observations on questionnaires filled out immediately after the street phase was finished. The questionnaires elicited information on conditions in the sites, how the census enumerators carried out the enumeration, and asked for the observers' estimates of the lowest number, and the highest number, of persons present at the site during the entire enumeration period. Each observer also filled out an Individual Census Report form ''as you believe the enumerator filled it out for you.'' District office personnel matched these dummy forms against the census forms to remove census forms for observers who were enumerated.

In order to ensure that census and observer results refer to the same sites, U.S. Census Bureau staff matched geographic information from the official census lists against geographic descriptions of sites from researchers' lists and observer questionnaires. Since a critical assessment variable is whether or not census enumerators were observed at the site, it was essential to determine whether during the designated time period observers were stationed at the sites the U.S. Census Bureau intended to enumerate.

Because enumerators were instructed to enumerate everyone visible at the site (except for people in uniform or engaged in obvious money-making activities), and to interview all who were awake, it was expected that every plant would be interviewed. Plants were instructed to remain in sight and allow themselves to be interviewed by enumerators. Thus, it is reasonable to take plants' enumeration experience as representative of other site occupants. The determination that a plant was captured during street enumeration

is based upon the plant's affirmative answers to either of two questions asked in the debriefing questionnaire: ''Were you interviewed by an enumerator?'' (Yes, No), and ''Do you think you were counted by an enumerator without being interviewed?'' (Yes, Maybe, No).

In this article, we use data from the assessment of street enumeration to explore the feasibility of a method which eventually may be used to obtain estimates of coverage of the street-dwelling homeless, and address some measurement problems involved in applying the method to street observers' reports. In the next section, we describe the statistical model, an alternative to the classical capture–recapture model which eliminates the need for the assumption of a closed population, and describe the assumptions underlying the model. Next, we describe the data on the basis of which the capture probabilities are estimated, and discuss their limitations and validity. Finally, we apply the model to data from all five study areas, and discuss some implications of our findings for future applications of this method.

2.2. *Statistical method*

There are two stages required to estimate the size of a closed population during a fixed time period using capture–recapture models (Seber 1982; Pollock 1991). In the most common version, in the first stage N_1 members of the population are captured, marked, and released. In the second stage, N_2 members are recaptured, and m, the number marked, is noted. The numbers captured in the first and second stages are random variables. Under the assumption of closure, and that the population at risk to be captured at the two stages is one and the same, the Lincoln-Peterson estimator of the total population is N_1N_2/m. For populations whose members are elusive, or for which the assumption of closure is questionable, this approach is both practically and theoretically unsatisfactory.

An alternative technique involves only one capture step. In this method, described by Laska and Meisner (1993), a known number of marked but otherwise indistinguishable members are ''planted'' among the population whose size is to be estimated. Although identifiable after capture, these plants are assumed to be intermixed in the population so that they have a capture probability equal to that of any other population member. As with the capture–recapture model, members of the augmented population are captured. The proportion of plants that are caught is an estimate of the probability of capture which, in turn, permits an estimate of population size.

Similar ideas can be identified in the literature. To determine software reliability, Duran and Wiorkowski (1981) formalized an unpublished method introduced by Mills, which involved seeding errors in computer codes. Earlier, Vilkitis (1968, 1971) utilized a technique he called violation simulation for estimating illegal big game kills.

The notation used in this application of the model is as follows:

H the unknown number of homeless individuals eligible for enumeration[4]

[4] Strictly speaking, not all individuals eligible for enumeration were homeless; anyone found in a street site between 2 and 4 a.m. March 21 who was not in uniform or engaged in money-making activities was eligible for enumeration. Theoretically, individuals who came to a designated site anytime during the 2–4 a.m. period, even for a few minutes, were eligible. In practice, individuals who visited a site but were not present while the enumeration was conducted would have been missed from the count of that site, and possibly missed entirely if they were not counted in some other site. It was also possible for individuals who visited more than one site to have been counted more than once. Ethnographic observations conducted in conjunction with this research suggest that mobility was significant in only a few sites, such as transportation terminals.

- R the known number of planted individuals who are indistinguishable from persons in street sites
- p the unknown capture probability that each individual (whether homeless or plant) is enumerated
- n_H the observed number of enumerated homeless
- n_R the observed number of enumerated plants
- n the sum of n_H and n_R is the total observed enumerated

The maximum likelihood estimator for H may be derived as follows: The joint distribution of the random variables n_H and n_R may be written as the probability of enumerating n individuals times the probability of enumerating n_H homeless. Given n individuals were enumerated, the likelihood is

$$L(H, p; n_H, n_R) = \binom{H+R}{n} p^n (1-p)^{H+R-n} \frac{\binom{H}{n_H}\binom{R}{n_R}}{\binom{H+R}{n}}$$

The first term is the binomial probability of obtaining n captures from among the $H + R$ individuals, and the second term is the hypergeometric distribution of having these n captures comprised of n_H homeless and n_R plants. Since $n = n_R + n_H$, L simplifies to a product of two independent binomials. Thus

$$L(H, p; n_H, n_R) = \binom{H}{n_H} p^{n_H} (1-p)^{H-n_H} \binom{R}{n_R} p^{n_R} (1-p)^{R-n_R}$$

For H known, the MLE of p is $[n/(H + R)]$, and for p known, the MLE of H is $[n_H/p]$ where $[x]$ denotes the greatest integer in x. Since the MLEs must simultaneously satisfy both of these expressions, solving for p and H yields:

$$\hat{p} = \frac{n_R}{R},$$

$$\hat{H} = \left[\frac{R n_H}{n_R}\right] = \left[\frac{n_H}{\hat{p}}\right]$$

The known value R is analogous to the number captured and marked in the first sample in a standard capture–recapture model, n_R corresponds to the number captured in both samples, and the MLE, $\hat{H}$, has the same form as the Lincoln-Peterson estimator mentioned above. Since there is a positive probability that $n_R = 0$ the estimator $\hat{H}$ does not admit a finite expectation.

Laska and Meisner (1993) proposed a nearly unbiased estimator to circumvent the problem of the infinite expectation of the MLE $\hat{H}$. The estimator, which is a hybrid of techniques proposed by Bailey (1951) and by Chapman (1951), is

$$\tilde{H} = (R+1)\frac{n_H}{n_R + 1}$$

The estimator $\tilde{H}$ has a finite expectation close to H, with

$$E[\tilde{H}] = H(1 - (1-p)^{R+1})$$

and variance

$$Var(\tilde{H}) = (R+1)H(1-p+pH)\beta(R+1;p) - H^2(1-(1-p)^{R+1})^2$$

where

$$\beta(R+1;p) = \sum_{j=1}^{R+1} \frac{1}{j} \binom{R+1}{j} p^j (1-p)^{R+1}$$

A confidence interval may easily be derived based on Fieller's theorem. See Laska and Meisner (1993) for further details on the derivation of the confidence interval for $\tilde{H}$ and other mathematical details.

2.3. *Assumptions of the method*

In order to apply the plant-capture method to data collected from street observers, several assumptions are required. The method assumes that (1) street observers' reports of their enumeration experiences are accurate, (2) the census enumerators carried out the planned survey as instructed and were readily observable by the plants while doing so, and (3) plants are ''well mixed'' into the total population (i.e., they are representative of the enumeration experiences of other occupants of the sites designated for census enumeration). Several factors could influence the validity of these assumptions.

2.3.1. Accuracy

Visual obstructions, or problems identifying site locations or boundaries, can reduce the reliability of observation. Hopper (1991) describes the barriers to observation posed by sites which were large and contained multiple structures, levels, and passageways. In such places, the identification of the boundaries of the designated site may have been problematic, and it was quite possible for enumerators to have visited parts of the site and remain unseen by observers stationed in other parts. In some such sites, enumerators may have come to a site and enumerated it (or some part of it) without being seen by the observers. This could lead to upward biases in the estimation of the number of homeless.

2.3.2. Enumeration

Unobtrusive enumeration. S-Night enumerators were trained to interview the inhabitants of the sites to which they were assigned, but were also permitted to enumerate by observation. For example, sleeping respondents were not to be wakened for an interview; rather, enumerators were instructed to estimate age, and to record race and sex based on observation. Based on observers' reports, it appears that some sites (especially in New York and Chicago), were enumerated entirely by observation. If enumeration was conducted unobtrusively, observers may not have realized they were enumerated, or indeed, may not have known that enumerators were present at all.

Timing of observation and enumeration. If census enumerators enumerated the site sometime other than 2–4 a.m., or if observers arrived late or left early, or left the site for any period of time, then the site may have been enumerated when observers were not present. If so, the plants' enumeration experiences would not be representative of the experiences of other site occupants. Of course, in this event the enumerators or the plants (or both) did not follow the original intended protocol.

For various reasons, then, observers' reports that they did not see enumerators may not mean that there were no enumerators present at the designated site. In general, the various barriers to observation make it more likely that observers would fail to see enumerators who were present than that they would report false sightings of enumerators when none were present. However, the latter was possible, and some observers may have mistakenly identified other people present in or driving by the site as census enumerators.

All of these factors would have resulted in an upward bias in the estimated number of homeless.

2.3.3. Mixing

The method should be applied by assigning the number of plants to be a constant proportion of a best guess of the number of homeless in each site. (Mixing plants at a uniform rate is intended to help meet the assumption of homogeneity, i.e., that plants should be effectively indistinguishable from the population to be counted. It is also intended to equalize across sites, and minimize, any effect of the observers on the enumeration itself.) In general, the number of plants assigned to a site in the S-Night assessment was larger when large numbers of persons were expected to be found there, but this was not carried out in a systematic or complete fashion. In the future, careful planning and strategies for assigning plants to sites should be included in the data collection design.

The model also requires the assumption that the capture probability of homeless people was not affected by the presence of the plants. Mathematically, this is equivalent to requiring that the capture probabilities of the homeless and of the plants are equal. If the plants are more (less) observable then the bias will be negative (positive). This corresponds to ''trap happiness'' (''shyness'') in the wildlife literature.

The assumption that plants' enumeration experience is representative of other site occupants would not be valid if other site occupants left when plants or enumerators arrived. Although we do not address the issue here, there were relatively few reports of homeless persons avoiding enumeration – for whatever reason – by leaving sites during the census.

3. Results

3.1. Estimating the probability of capture

Ultimately, 16 of 156 observer sites could not be matched to the census. The 16 unmatched sites are excluded from analysis, leaving 140 sites. (For more details about matching see Martin 1992.) Official census counts were returned for 130 of the 140 sample sites, with a total of 1,803 people counted, as shown in Table 1. Ten sites, all in Los Angeles or Phoenix, were eliminated at some stage in the census process and final counts were not processed through the official census count. (These sites, which contained a total census count of 49, appear to have been eliminated as a result of local review, recanvass, or close-out activities occurring at the time, for reasons which are unknown.) These sites are not shown in Table 1.

As shown in Table 1, the number of people counted in the sample sites varies from 23 in the Chicago study area to 1,318 in New York. Part of this variability occurs because the

Table 1. Total population counted at matched, census-defined sites, by study area

	New Orleans	New York	Phoenix	Los Angeles	Chicago	Total
Census count	109	1,318	135	218	23	1,803
Number of sites	18	54	21	23	14	(130)

Table 2. Per cent of observers who report being interviewed or counted, by study area

	New Orleans	New York	Phoenix	Los Angeles	Chicago
Interviewed	67	37	44	33	7
Not interviewed					
Counted	10	17	8	2	0
Maybe counted	7	12	3	4	18
Not counted	10	20	10	13	25
Did not see enumerators	5	14	36	48	50
Total per cent	100	100	100	100	100
Number of observers	58	104	39	46	28

Note: Per cents may not sum to 100 due to rounding error. Sixteen sites in which positive census counts were returned but no enumerators were seen are excluded. (See text for discussion of these sites.)

New York study area covers four U.S. Census Bureau District Office areas, while each of the other study areas covers just one district. There are additional, unknown variations among the areas in the size of the street population available for enumeration, as well as variations among areas in the completeness of census coverage of the street population eligible for enumeration. Census coverage was likely affected in varying degrees by how appropriate and complete the set of sites selected for enumeration was, by how completely the enumeration was conducted by the district offices and census enumerators, and by the behavior of homeless and other persons (including the news media) at the sites.

Table 2 shows that the proportion of plants who report they were interviewed ranged from two-thirds in New Orleans to only seven per cent in Chicago. An additional 6 to 29 per cent in each study area said they were not interviewed but were, or may have been, counted. Plants in New York and Chicago were most likely to express uncertainty about whether they had been counted. These were areas where census enumerators more commonly counted site occupants by observation and did not conduct interviews with them. Between 5 (in New Orleans) and 50 per cent (in Chicago) of observers reported never seeing census enumerators in the sites where they were stationed.

As discussed above, various barriers to observation may have affected the reliability of the reports of plants who said they did not see enumerators. One indicator of the validity of observers' reports is their degree of agreement with census outcomes. Table 3 presents the cross-classification of observer reports and census outcomes for 128 sample sites,[5] by study area. If any observer at a site saw an enumerator, or if enumerators were seen before or after the scheduled enumeration period of 2–4 a.m., or were seen near a site but not in it, these reports were counted as positive reports of enumerator presence. In interpreting this table, we take a positive census count as indicating that census enumerators came to a site

[5] For 2 of the 130 sites, enumerator presence could not be ascertained.

Table 3. Consistency of observer reports and census outcomes, by study area

	Was there a positive census count?	Were enumerators seen?		
		Yes	No	Total
New Orleans	Yes	7	0	7
	No	11	0	11
	Total	18	0	18
New York – S. Manhattan	Yes	12	10	22
	No	1	2	3
	Total	13	12	25
New York – Other District Offices	Yes	16	0	16
	No	9	3	12
	Total	25	3	28
Phoenix	Yes	7	2	9
	No	5	7	12
	Total	12	9	21
Los Angeles	Yes	9	1	10
	No	2	11	13
	Total	11	12	23
Chicago	Yes	4	3	7
	No	0	6	6
	Total	4	9	13

and counted the people there. A zero count could occur either if census enumerators came to a site and there was no one to count, or if they failed to enumerate a site. Thus, a clearly inconsistent joint outcome occurs when observers say they saw no enumerators, yet there is a positive census count for a site. These data show that census outcomes and observer reports are mostly but not wholly consistent. Overall, there are positive census counts in 66 per cent (55 of 83) of sites where enumerators were seen, compared to 36 per cent (16 of 45) of sites where enumerators were not seen. Conversely, enumerators were seen in 77 per cent (55 of 71) of sites with positive census counts. There are 16 anomalous sites in which census counts are positive yet enumerators were not seen; ten of these occurred in one District Office in South Manhattan. If the results for the South Manhattan District Office are excluded, census and observer results are quite consistent: enumerators were observed in almost 90 per cent of sites which had positive census counts, and census counts of 0 were returned in 82 per cent of sites where no enumerators were observed.

As discussed above, the discrepancies between observer reports and census outcomes may be due to many factors. Observers may not have been at the site at the same time as the enumerators, or may have visited different portions of a site. Curbstoning by enumerators could result in positive counts for sites at which enumerators were not observed. Site visits and investigation of the U.S. Census Bureau records for the discrepant South Manhattan sites suggested that most or all of them were visited, but the observer failed to see the enumerator. There were some timing problems, some drive-by enumerations may have

occurred, and several sites also had ambiguous boundaries or their exact locations were not clear from the address descriptions given in census materials, so enumerators and observers may have been in slightly different places (Schwede 1991).[6]

Because an observer's report of not seeing census enumerators is ambiguous in its implications, probability of capture was estimated on the basis of data from sites where enumerators were seen by observers. However, this tends to over-estimate the probability of capture. For many sites, there is no reason to doubt the observers' reports that enumerators failed to appear. Further, most of the sites where enumerators were not seen returned census counts of zero (see Table 3). If, however, enumerators came outside of the designated time frame, they may have obtained valid counts but the plants had not yet arrived or had already left. This obviously violates the assumption that each plant and each homeless site occupant are captured with the same probability. Ultimately, any actual application of this method would require more reliable observational data, or a way to distinguish valid and invalid negative reports.

3.2. *Applying the model to estimate capture probabilities*

As described above, a nearly unbiased estimator of the capture probability p is provided by $(n_R + 1)/(R + 1)$. Below, we use two alternative measures of n_R, the number of observers who were counted: n_{R1} does not include the ''maybes,'' while n_{R2} does. These, in turn, give rise to $\hat{p}_1$ and $\hat{p}_2$, with $\hat{p}_1 < \hat{p}_2$.

Table 4 presents estimated capture probabilities for the five study areas, conditional on enumerators being seen in a site ($\hat{p}_{\text{seen}}$ denotes the estimated probability that enumerators were seen at the site, $\hat{p}_i$, $i = 1, 2$ denotes the estimated conditional capture probabilities, and R_{seen} is the number of plants in sites where enumerators were seen).

The values of $\hat{p}_{\text{seen}}$, the estimated probability that enumerators were seen by observers, are highly variable among study sites, ranging from .31 in Chicago to 1.0 in New Orleans. Although $\hat{p}_{\text{seen}}$ is a fallible measure of the probability that sites were visited (as we have discussed, some sites appear to have been enumerated without the observers' knowledge), these values suggest that sites were missed with greatly varying estimated probabilities in the different study areas. On the other hand, given that enumerators were seen, the estimated capture probabilities are fairly high and uniform among study areas, with the exception of Chicago. The overall estimated capture probabilities of .77 (for the

Table 4. Capture probabilities, conditional on enumerators being seen

City	R_{seen}	$\hat{p}_{\text{seen}}$	$\hat{p}_1$	$\hat{p}_2$
Chicago	13	.31	.21	.57
New Orleans	58	1.0	.78	.86
Phoenix	26	.57	.78	.81
New York	94	.72	.60	.74
Los Angeles	25	.48	.65	.77
Total	216	.65	.65	.77

[6] Ironically, the census counts and observer counts aggregated for these sites are highly consistent (see Martin 1992), suggesting that both enumerators and observers enumerated the site accurately even though the observers did not see the enumerators.

Table 5. Population counts obtained by census, estimated by observers, and projected by the model for sites where enumerators were seen

City	Number of sites	Census counts	Observer estimates of number of people in sites		Projected population size ($\tilde{H}$) based on model	
			Mean low numbers	Mean high numbers	Based on n_{R2}	Based on n_{R1}
Chicago	4	11	18	50	19	51
New Orleans	18	109	118	242	126	140
Phoenix	12	104	110	140	128	134
New York	38	1,240	551	900	1,683	2,067
Los Angeles	11	217	196	250	282	332
Total	83	1,681	993	1,582	2,238	2,724

more inclusive $\hat{p}_2$) and .65 (for $\hat{p}_1$) suggest that enumeration was fairly complete for sites that were visited. Estimates below 1.0 are consistent with observers' reports that enumerators approached people selectively within many sites. There was widespread publicity surrounding S-Night as a ''count of the homeless,'' and this is how many census enumerators understood their task. Despite the U.S. Census Bureau's emphasis in training on the importance of counting everyone in the sites, some enumerators ignored the procedure and improvised their own ways of ''counting the homeless'' (for an xample, see Martin 1990). Overall, 19 per cent of observers report that enumerators approached only people who appeared homeless, with no statistically significant variation among study areas. Additional reasons for less-than-complete enumeration include insufficient time to complete the enumeration, as well as other types of problems at the sites (e.g., drug dealing), as reported by observers.

Table 5 presents several sets of population counts for sites where census enumerators were seen. Census counts are compared with counts estimated by observers, and counts projected on the basis of the model. The observer estimates, say, for the low numbers, represent means of these counts, averaged over observers within sites, and summed across sites within study areas.[7] Two sets of population projections are presented in the last two columns of Table 5. The count n_{R2}, which includes the ''maybes,'' yields higher capture probabilities (see Table 4, above) and therefore lower population projections, than n_{R1}.

For four of the five study areas, the three sets of counts are reasonably consistent, with census counts at the low end of both the observer estimates and the model projections. In general, the model projections are fairly consistent with the observer estimates. The exception is New York, where the census counts exceed the observer estimates by large margins, and the model projections are even higher than the census counts. In some New York sites, lack of comparability between sites as defined by the census and by observers implies that the two sets of counts refer to different entities. The most prominent example is a large transportation terminal where 653 people were counted by teams of enumerators who moved throughout the terminal. Observers were stationed at specific locations within

[7] For several sites in New Orleans, Chicago, and New York, observer estimates were summed rather than averaged. This occurred when observers were stationed at zones within the site, so their counts referred to only part of a site.

Table 6. Capture probabilities for ''easy'' sites, by study area

City	R_{seen}	$\hat{p}_1$	$\hat{p}_2$
Chicago	1		
New Orleans	30	.74	.87
Phoenix	17	1.0	1.0
New York	25	.77	.86
Los Angeles	2		
Total – ''easy'' sites	75	.80	.87
Total – ''difficult'' sites	141	.56	.72

the terminal, and their counts referring to those areas sum to 216 (low) and 348 (high). This problem, and particularly this site, contribute to low counts by New York observers relative to census.

In order to further investigate the plausibility of the estimates produced by the model, we redid the analysis eliminating difficult sites in which there was reason to believe that the site itself was problematic in some way, or that enumerators had carried out the enumeration selectively.[8]

If conditions are conducive to enumeration as well as observation, and *if* enumerators follow the procedure of enumerating everyone, then capture probabilities should begin to approach 1. These results are shown in Table 6. R represents the number of plants in easy sites where enumerators were seen. Since R is so small in Chicago and Los Angeles, capture probabilities are not given for those study areas. Table 6 shows that the capture probabilities for the other three study areas are uniform and high, as hypothesized. For comparison, capture probabilities are also presented for the total difficult sites, and they are much lower. Note also that the alternative estimates of capture probability cover a wider range for the stratum of difficult sites than for the stratum of easy sites (the ratios of $\hat{p}_2$ to $\hat{p}_1$ are 1.29 and 1.09, respectively).

Table 7, in contrast to Table 5, presents counts for these easy sites (again only including sites where enumerators were seen). These results suggest that in the easy sites in two study areas (New York and New Orleans), the census counts are within the bounds of the low and high observer estimates. In Phoenix, census counts exceed the high observer estimates. In Los Angeles and Chicago, easy sites are not available in sufficient numbers to permit comparisons. Consistent counts would be expected to the extent that there are no barriers to enumeration or observation, and enumerators approach people nonselectively; this expectation is confirmed in only two study areas. In addition, the site population sizes projected by the two alternative capture probabilities cover a narrower range (279–299, compared to 2,238–2,724 based on Table 5), as would be expected when observers were better able to determine if they actually had been enumerated or not.

[8] We tallied observer reports of any types of problems at the site (e.g., visual barriers, drug activities, observer was thrown out of or harassed at site, heard gunshots) which might have disrupted either enumeration or observation; and reports that enumerators did not approach everyone or did not have enough time to complete the enumeration. If more than 50 per cent of observers mentioned any of these problems, the site was classified as ''difficult''; otherwise, it was classified as ''easy.''

Table 7. Population counts for easy sites

City	Number of easy sites	Census counts	Observer estimates of number of people in sites		Projected population size ($\tilde{H}$) based on model	
			Mean low numbers	Mean high numbers	Based on n_{R2}	Based on n_{R1}
Chicago	1	6	3	13	12	12
New Orleans	9	76	47	81	87	102
Phoenix	8	95	69	72	95	95
New York	12	69	54	83	85	90
Los Angeles	1	0	1	6	0	0
Total	31	246	174	255	279	299

4. Conclusions

Our exploratory analysis suggests that the plant-capture method proposed by Laska and Meisner (1993) holds promise for estimating coverage of elusive populations, such as the street homeless, for which the traditional capture–recapture method is inadequate. The essential requirements for applying this method are that sites must be well-identified with clear boundaries, the space within them must be observable, enumerators must conduct the enumeration in a timely fashion, and they must be readily observable. Since the method relies on observational data, it is vulnerable to environmental conditions and departures from prescribed procedure that affect the observability of enumeration. As our data suggest, conditions conducive to observation did not always exist in the street sites enumerated in the 1990 census.

The plant-capture method could encounter fewer problems if enumerations were to be conducted in sites or places with clear boundaries. For example, another method of enumerating street homeless is to conduct daytime counts in soup kitchens and other service facilities. The plant-capture method would seem well-suited to this situation, and could yield valuable information about the completeness of enumeration. Additionally, the plant-capture method has promise as a method for estimating coverage within homeless shelters and some other types of group quarters in which occupants are actually interviewed, and for which the U.S. Census Bureau has very little information about within-facility coverage. If reliance were to be placed on observer reports as a basis for estimating coverage of these or other places, then observers would probably need more extensive training in observational techniques than was given to the 1990 S-Night observers.

A key determinant of the coverage of the street homeless in the 1990 census was the completeness of the set of sites targeted for enumeration. The U.S. Census Bureau identified sites in advance by inviting cities, advocates, and others to submit lists of public sites where homeless people congregate at night; additional sites were added by district office personnel.[9] However, there is no way of knowing how well this list represented the set of public places where street homeless people could have been found between 2 and 4 a.m. March 21, 1990. As it stands now, the plant-capture method discussed here leaves

[9] For the most part, the U.S. Census Bureau did not enumerate ''hidden homeless'' who were not in public places, although there was an attempt to count persons emerging from abandoned buildings the morning of March 21.

unanswered questions regarding the coverage of the frame of street sites. In order to develop a full measure of coverage of street homeless, a way of evaluating the completeness of the sampling frame would need to be developed. Other approaches to enumerating street homeless which involve different types of frames are more conducive to evaluation of frame coverage (e.g., the area sample used by Rossi et al. 1987, or the list of soup kitchens and service facilities used in the 1987 HUD survey and in the U.S. Census Bureau's 1989 pilot test in Baltimore by Campanelli et al. 1990), or an adaptation of statistical techniques designed to estimate the number of species in a region could be used to estimate the number of unlocated street sites; see Seber 1982.

5. References

Bailey (1951). On Estimating the Size of Mobile Populations from Capture-recapture Data. Biometrika, 38, 293–306.

Campanelli, P.C., Salo, M.T., Schwede, L., and Martin, E. (1990). Research on Enumerating the Homeless: Results of a U.S. Census Bureau Test of Alternative Methods. Paper presented at the Meeting of the American Statistical Association, Anaheim, CA.

Chapman (1951).

Duran, J.W. and Wiorkowski, J. (1981). Capture–Recapture Sampling for Estimating Software Error Content. IEEE Transactions in Software Engineering, Vol. SE-7, No. 1, January.

Hopper, K. (1991). Repeat Enumerations, Structured Interview, and Brief Ethnographic Studies of the Street Census Project. Final report for JSA 90-19 between the U.S. Census Bureau and Nathan Kline Inst., Orangeburg, NY, May.

Laska, E.M. and Meisner, M. (1993). A Plant-Capture Method for Estimating the Size of a Population from a Single Sample. Biometrics 49, 209–220.

Martin, E.A. (1990). Trip Report – S-Night in Los Angeles. U.S. Census Bureau Memorandum, May 30.

Martin, E.A. (1992). Preliminary Findings of Assessment of S-Night Enumeration, 1990 Decennial Census. Preliminary Research and Evaluation Memorandum, No. 112, Washington, D.C.: U.S. Bureau of the Census.

Pollock, K.H. (1991). Modelling Capture, Recapture and Removal Statistics for Estimation of Demographic Parameters for Fish and Wildlife Populations: Past, Present, and Future. Journal of the American Statistical Association, 86, 225–238.

Rossi, P.H., Wright, J.D., Fisher, G.A., and Willis, G. (1987). The Urban Homeless: Estimating Composition and Size. Science, 235, 1336–1341.

Schwede, L. (1991). Investigation of Reasons for Anomalies in DO 2201 Street Sites. U.S. Census Bureau, November 25.

Seber, G.A.F. (1982). The Estimation of Animal Abundance and Related Parameters. New York: Macmillan.

Taeuber, C.M. and Siegel, P. (1991). Counting the Nation's Homeless Population in the 1990 Census. In C.M. Taeuber (ed.) Enumerating Homeless Persons: Methods and Data Needs. Washington, D.C.: U.S. Census Bureau.

Vilkitis, J.R. (1968). Characteristics of Big Game Violators and Extent of Their Activity in Idaho. M.S. Thesis, University of Idaho, Moscow.

Vilkitis, J.R. (1971). The Violation Simulation Formula Proves Reliable as Field Research in Estimated Closed-Season Illegal Big Game Kill in Maine. Transactions of the Northeast Section of the Wildlife Society, Vol. 28, 141–144.

Received December 1994
Revised April 1996