Enumerating the Hidden Homeless: Strategies to Estimate the Homeless Gone Missing From a Point-in-Time Count

Robert P. Agans[1], *Malcolm T. Jefferson*[1], *James M. Bowling*[1], *Donglin Zeng*[1], *Jenny Yang*[1], *and Mark Silverbush*[2]

To receive federal homeless funds, communities are required to produce statistically reliable, unduplicated counts or estimates of homeless persons in sheltered and unsheltered locations during a one-night period (within the last ten days of January) called a point-in-time (PIT) count. In Los Angeles, a general population telephone survey was implemented to estimate the number of unsheltered homeless adults who are hidden from view during the PIT count. Two estimation approaches were investigated: i) the number of homeless persons identified as living on private property, which employed a conventional household weight for the estimated total (Horvitz-Thompson approach); and ii) the number of homeless persons identified as living on a neighbor's property, which employed an additional adjustment derived from the size of the neighborhood network to estimate the total (multiplicity-based approach). This article compares the results of these two methods and discusses the implications therein.

Key words: Homeless count; hidden homeless; unsheltered homeless population; Horvitz-Thompson estimator; multiplicity-based estimator.

1. Introduction

How to best estimate homelessness has historically been a difficult and costly venture (Link et al. 1994; Tompsett and Toro 2004; Toro and Janisse 2004; Toro 2005, 2006). As a highly mobile population, it is difficult to contact and track the homeless due to their unstable living situations. Furthermore, when performing homeless counts of the unsheltered population, enumerators are typically required to count late at night when the shelters are closed for the evening and the homeless on the streets are more easily identifiable. Night-time counts, however, can be problematic because visibility is reduced and vulnerable populations, such as children, hide from public view. In addition, homelessness is not a permanent state. A person's housing situation can change rapidly and homeless people can relocate quite easily. Thus the true value of homelessness is constantly in flux, which creates inherent variability between estimates taken at different points in time. Estimates therefore can vary depending on the assumptions and

[1] University of North Carolina, Carolina Survey Research Laboratory, 730 Martin Luther King Jr. Blvd., Bolin Creek Center, Chapel Hill, NC, U.S.A. Corresponding author Email: agans@unc.edu
[2] Los Angeles Homeless Services Authority, 811 Wilshare Blvd., Los Angeles, CA 90017, U.S.A

methodology applied in the count. Nevertheless, homelessness needs to be measured, especially in the U.S., which has one of the highest rates of homelessness among developed nations (Tompsett et al. 2003; Toro et al. 2007). In addition to the important social issues that surround homelessness, there are practical reasons to obtain the best possible estimates. One such agency that depends on reliable measures of the homeless population is the Los Angeles Homeless Services Authority (LAHSA).

LAHSA, a joint powers authority of the County and City of Los Angeles, coordinates and manages government funds for programs that provide shelter, housing, and other services to the homeless in 85 of the 88 cities of Los Angeles County (and all of the unincorporated areas). In order to receive federal dollars, LAHSA is required to conduct a homeless count every two years. Given that Los Angeles is the largest urban county in the U.S. with more than ten million residents and a geographic area of 4,083 square miles that encompasses 88 cities, this task is challenging. Los Angeles County also has one of the largest disparities between wealthy and low-income people in the nation. It manages one of the largest welfare systems in the country and contends with one of the nation's largest homeless populations (Bring Los Angeles Home 2006).

In 2009, the Carolina Survey Research Laboratory (CSRL) at the University of North Carolina at Chapel Hill collaborated with LAHSA for the 2009 Los Angeles Homeless Count (HC09). Homelessness was measured at the Continuum of Care (CoC) level and included all of Los Angeles County except the cities of Glendale, Pasadena, and Long Beach, which produce their own independent estimates. HC09 included a street count, a shelter count, and a youth count as well as a hidden homeless estimate, all in an attempt to measure the county's homeless population. The HC09 findings estimated that 42,694 people were homeless and that over 20 percent of the homeless population was hidden (2009 Greater Los Angeles Homeless Count Report). The hidden homeless estimate was derived from interviewing 4,288 households in the CoC and asking them if a hidden homeless person lived on their property. Though a relatively large number of sample respondents provided interviews, only 16 hidden homeless persons were identified as meeting the HUD criteria for hidden homelessness. The rarity of this event produced a rather imprecise estimate with a relatively large variance where the total was 9,451 and the standard error 2,339.

The focus of this article is on the hidden homeless estimate, which made up a significant portion of the homeless population in HC09. To redress the issue of large sampling errors commonly associated with survey estimates of rarely occurring events, frame coverage was increased in the 2011 Los Angeles Homeless Count (HC11) by asking survey participants to report not only on homeless individuals currently residing on private property, but also on homeless individuals in their immediate neighborhood. Though this method increases the number of hidden homeless persons that get reported, it also introduces the potential for multiple reports of the same hidden homeless persons by members of the same neighborhood network. Consequently, a multiplicity-based estimator must be derived which adjusts for the possibility of multiple reports of hidden homeless. It is through this neighborhood network approach that we hope to improve the statistical precision of the hidden homeless estimate. This article examines the utility of the multiplicity-based approach for estimating the hidden homeless population in Los Angeles.

2. Methods

To minimize the difficulties that arise in sample surveys designed to estimate a rare event, Birnbaum and Sirken (1965) proposed a stratified random sample design that included obtaining information about the "multiplicity" of individuals with a rare condition as reported by health care providers (i.e., the primary sampling units). To derive an estimate of the number of diagnosed cases of a rare disease, these investigators designed the survey to increase sample coverage while accounting for patients being reported by several providers. Following this idea, an alternate approach was proposed in HC11 that expands beyond reporting for hidden homeless individuals located on a private property, as was the sole approach for HC09. In the present study, respondents were asked to not only report the number of hidden homeless on their residential property but also any hidden homeless on their neighbor's property. Consequently, coverage was expanded and the possibility of discovering hidden homelessness was broadened, thus hopefully making a rare event less rare. This approach has been called network or multiplicity sampling in the literature (Sudman and Freeman 1988; Flores-Cervantes and Kalton 2008; Kalton 2009), because the typical one-to-one counting rule in conventional sampling takes into account the inclusion of larger observational units or networks. Instead of being self-weighting (conventional approach), the multiplicity approach must take into account the network size. The success of this approach, however, rests largely on the validity and accuracy of the information provided by the respondent, not only for the critical measure of hidden homelessness, but also for a new, additional measure of neighborhood network size. In order for the multiplicity estimate to be valid, we must accurately gather the size of the neighborhood and the probability that more than one person (viz., neighbors) could report on the same hidden homeless person or persons. In this article, comparisons were made between the statistical precision of a more conventional Horvitz-Thompson approach (as applied in HC09) and the multiplicity-based approach proposed here.

2.1. Sample Design

Households for the telephone interview were identified from a disproportionately stratified dual-frame sample of landline telephone numbers obtained from Marketing Systems Group in Horsham, Pennsylvania. Stratification was defined by frame source (list-assisted RDD and electronic white page "EWP" listings), median household income of the exchange area (EA) in which the telephone number was located (delineated into *high* and *low* designations at $50,680), the percent of single-family households in the EA (delineated into "high" and "low" designations at 60%), as well as other local area information thought to be predictive of hidden homelessness (Table 1). The latter was used to form an *item predictor score* (*IPS*) based on the distribution of such scores for all listed telephone numbers on the frame. IPS was considered *low* in designated areas if it resulted in a summative score of 0 or 1, and high if the IPS ranged between 2 and 6. IPS scoring criteria consisted of the following items:

- Single family dwelling: Yes = 1; No = 0;
- High African American concentration: Yes = 1; No = 0;

Table 1.   Strata used to sample telephone numbers for the hidden homeless telephone survey

			Median Household Income in Exchange Area:			
			Low		High	
			Listing Status of Phone Number:		Listing Status of Phone Number:	
			Directory Listed	NOT Directory Listed	Directory Listed	NOT Directory Listed
% Single Family Dwelling Units in Exchange Area:	High — Item Predictor Sum:	High	1	5	2	7
		Low	9		11	
	Low — Item Predictor Sum:	High	3	6	4	8
		Low	10		12	

- Below the 20th percentile for length of time in current residence (measure of mobility): Yes = 1; No = 0;
- Below the 20th percentile on household income: Yes = 1; No = 0;
- In a block group (or census tract, if only available at this level) above the 80th percentile on percent vacancy rate: Yes = 1; No = 0; and
- In a Census Tract above the 80th percentile on rate of street homeless count per 100,000 population members as of the 2000 Census: Yes = 1; No = 0.

2.2.   Questionnaire Design and Pilot Testing

In general, *hidden homeless persons* are defined as those who live among, but not directly with, the residential population of a community. In this study, people were classified as hidden homeless if they were sleeping on private property in such places as an unconverted garage, carport, back porch, tool shed, tent, camper, car, encampment, and so on. These people were likely to go missing during a point-in-time count because they were not on the streets nor readily visible to enumerators, but nevertheless were considered unsheltered homeless according to the U.S. Department of Housing and Urban Development (HUD) definition. This definition is in contrast to the *precariously housed* or *at risk of literal homelessness* definitions, which include individuals temporarily staying within a household because they have no regular or adequate place to stay and lack the means or money to provide it.

Neighborhood was operationally defined by street block neighborhood (SBN). The SBN for any household was defined as the set of households whose front entrance faces the street – bounded by one linear segment of street on which the referent household is located. In other words, a respondent reported on the number of hidden homeless for the set of households on both sides of a one-block long street that also includes the respondent's household. Residents of apartment, condominium, or single room housing

complexes were asked to provide estimates of hidden homeless on complex property. An additional requirement was that respondents must also give a measure of neighborhood network size. For single family dwellings, respondents need to provide a count of homes contained within the SBN. For complexes, respondents need to report the number of housing units in their building and the number of buildings in their housing development in order to calculate their respective neighborhood network size.

For the pilot study, 2,500 random digit dial telephone numbers within the Los Angeles CoC were placed into calling between December 14 and December 29, 2010. Of those telephone numbers, 1,662 were finalized as ineligible, 55 as refusals, 747 were given an unknown status because eligibility was never determined, and 36 cases resulted in completed interviews. Interviews were completed with an adult who either owned or rented the property. After reviewing the pilot data, minor changes were made to the instrument to improve the flow of the telephone interview. The question about any hidden homeless on personal property read: *Not including dependents or adult children, is there anyone living with you or staying on your property because they do not have a regular or adequate place to stay due to a lack of money or other means of support?*

The questions on neighborhood hidden homeless piloted well, but went through minor reconstruction to simplify telephone administration. The previous version read: *Now I'd like to ask you some questions about your neighborhood. Have you seen anyone staying [on your neighbors' property (IF X10 = single dwelling)] [in your complex (IF X10 = housing complex)] who you believe does not have a regular or adequate place to stay due to a lack of money of other means of support? For this question, your [neighborhood includes houses on both sides of the street confined by two intersection streets] OR [complex includes all the housing units in your complex or development]. INTERVIEWER NOTE: THIS DOES NOT INCLUDE DEPENDENTS OR ADULT CHILDREN.*

The new version read: *Now I'd like to ask you some questions about the houses on your block. Have you seen anyone who appears to be homeless staying [on your neighbors' property (IF X10 = single dwelling)] [on the complex grounds (IF X10 = housing complex)] [WHO YOU BELIEVE DOES NOT HAVE A REGULAR OR ADEQUATE PLACE TO STAY DUE TO A LACK OF MONEY OF OTHER MEANS OF SUPPORT]? For this question, your [neighborhood block includes houses on both sides of the street confined by the closest intersection in either direction] OR [complex includes all the housing units in your complex or development].*

2.3. Main Study

All totaled, 32,826 randomly selected telephone numbers were worked to complete 3,390 hidden homeless interviews. The overall selected sample of telephone numbers was disproportionately allocated among the twelve strata, following a similar allocation pattern of disproportionality to that observed in the HC09 survey. Data collection took place from January 16, 2011 to April 10, 2011. The CSRL has an advanced CATI operation consisting of 42 interviewer workstations and three monitoring stations. Supervisors and clients can silently monitor interviewers' audio and keyed responses from computers in the monitoring room. This monitoring capability helps ensure that data

collection for the study meets the highest quality standards. During data collection, interviewing took place Saturday through Thursday (EST). Monday through Thursday shifts typically were conducted from 12 noon to 12 midnight. Saturday sessions occurred between 1:30 pm until 5:30 pm. Sunday shifts were typically held from 5:30 pm to 12 midnight.

In addition to questionnaire programming, the CSRL also utilizes Blaise's (Version 4.8, 2008) call-scheduling capabilities to maximize the probability of contacting potential respondents. A central file server takes sample telephone numbers and arranges automatic call scheduling for interviewer administration. The system enables calls to be scheduled so that different times of the day and week are represented. In this study, no cases were withdrawn from calling until a minimum of eight unsuccessful call attempts were made and had been at least one weekend call, one evening call, and one daytime call. Calls could also be scheduled at times specified by the respondent, thus ensuring that calls were made at optimum times.

CSRL supervisors closely monitor data collection to ensure that data are being collected and entered correctly according to guidelines and policies reviewed in training. In addition, several steps were taken to both reduce the occurrence of refusals and to improve refusal conversion. First, we attempted to minimize refusals by introducing techniques for dealing with reluctance and refusal during general interviewer training. This was often accomplished through role-playing sessions that enable trainees to become familiar with and to rehearse a variety of refusal situations. Upon encountering a refusal, interviewers documented the following information for each refusal: reason for the refusal, the point in the interview at which the refusal occurred, and the gender and approximate age of the respondent. Refusal documentation is standard procedure at the CSRL because it enables the next interviewer, the refusal converter, to tailor her approach in eliciting participation from the potential respondent, thereby optimizing the likelihood of conversion. Finally, as part of interviewer monitoring, interviewers' individual refusal rates were closely watched. Only experienced refusal converters recontact respondents who initially refuse.

2.4. Final Outcomes and Response Rates

The final outcomes from calling may be grouped into four broad categories (see Table 2) that were used to calculate the overall response rate: (i) a complete interview (I = 3,390); (ii) not eligible (NE = 13,503) because the telephone numbers were found to be nonworking, dedicated fax or computer lines, or a business/cell line; (iii) no interview or response from an eligible household (NR = 2,593); or (iv) unknown or indeterminate

Table 2.   Final dispositions for hidden homeless survey by strata

OUTCOME	STRATA												Totals
	1	2	3	4	5	6	7	8	9	10	11	12	
I	590	379	608	209	151	39	85	88	440	254	316	231	3,390
NR	518	357	467	94	120	32	50	48	378	190	192	147	2,593
NE	1,147	859	838	289	3,150	1,056	1,670	2,795	662	441	271	324	13,503
U	1,983	1,563	2,047	685	1,085	322	532	735	1,497	916	1,027	948	13,340

(U = 13,340) because we never had the opportunity to talk to a person or someone in the household refused participation before we could verify eligibility.

The response rate was based on the American Association for Public Opinion Research (AAPOR) Standard Definitions (2011). Weighted (42.6%) and unweighted (33.6%) response rates were determined by AAPOR's Response Rate 3 (RR3).

2.5. Final Sample Weights

A standard three-step sample weighting procedure was followed in producing sample weights (Kalsbeek and Agans 2008). The base weight was computed using the sampling rate for telephone numbers in each stratum, adjusted for the portion of the stratum samples that were placed in calling and the number of phone lines that reached the household (Step 1). The base weight was then adjusted for differential household-level nonresponse among sampling strata using the inverse of the stratum-specific household-level RR3 response rate as the adjustment factor (Step 2). The nonresponse-adjusted household sample weight was then calibrated to population counts as estimated from the American Community Survey sample by cross-tabulating on: (i) the race-ethnicity of the reference person/knowledgeable adult (white non-Hispanic/Hispanic/Other), (ii) the type of dwelling (single-family/all other types), and (iii) the education of the reference person/ knowledgeable adult (< college bachelor's degree/≥ college bachelor's degree) (Step 3). The multiplicative effect of variable sample weights or $Meff_w$ (Kish 1965) was somewhat large (Meff = 2.44), so weights were trimmed at the nonresponse adjustment stage as recommended by Potter (1988).

2.6. Horvitz-Thompson Estimator

The Horvitz-Thompson (HT) estimator of a population total was used to estimate the number of hidden homeless on private property for the HC11 survey. This conventional method was also employed to estimate the total number of hidden homeless individuals in the LA for the HC09 survey. First, we define

N	Number of closed street block neighborhood (SBN) networks in the target population
Y_i	The actual number of hidden homeless persons in the i-th SBN
M_i	Number of household residences in the i-th SBN
I_{ij}	Sample selection indicator for the j-th household in the i-th SBN $\Rightarrow$ 1 if household is selected, 0 otherwise
$\pi_{ij} = E(I_{ij})$	The selection probability for the j-th survey household in the i-th SBN

Now let t_{HH} denote the total (relevant) hidden homeless count that is to be estimated such that

$$t_{HH} = \sum_{i=1}^{N} Y_i = \sum_{i=1}^{N} \sum_{j=1}^{Mi} H_{ij}^{(Res)},$$

where $H_{ij}^{(Res)}$ denotes the number of hidden homeless persons located on their private property at the j-th survey household in the i-th SBN. Then the HT estimator of a

population total for the number of hidden homeless people identified in the HC11 survey is given as:

$$\hat{t}_{HT} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \frac{I_{ij} H_{ij}^{(Res)}}{\pi_{ij}}.$$

2.7. Multiplicity-Based Estimator

To apply the multiplicity-based alternative approach, the telephone survey interview required all respondents (located in the j-th household in the i-th SBN) to provide the number of hidden homeless persons located on the property of all other surrounding households in their SBN, also denoted as $H_{ij}^{(SBN)}$ Respondents who did not provide SBN size data were called back for data retrieval. Hot deck imputation was used to handle missingness for households still missing SBN size data after unsuccessful attempts to retrieve the information through a callback. Specifically, records for which measure of size data was missing were imputed using the mean SBN value of other record that shared that particular record's household type: For instance, in the case where a participant informed the interviewer that they lived in an apartment but refused or were unable to provide SBN size data, missing information was imputed using completed SBN data collected from the other participants who also classified their residence as an apartment.

Following the Birnbaum and Sirken (1965) approach to multiplicity estimation, an estimator of the overall number of hidden homeless persons in the target population is given as

$$\hat{t}_{HH} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \frac{I_{ij} Y_{ij}}{\pi_{ij} M_i}, \quad \text{where} \quad Y_{ij} = \left[H_{ij}^{(Res)} + H_{ij}^{(SBN)} \right].$$

Ignoring the biasing effects of nonsampling error (i.e., due to frame, nonresponse, and measurement), $\hat{t}_{HH}$ can be shown to be an unbiased estimator of t_{HH}:

$$E(\hat{t}_{HH}) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left\{ \frac{Y_{ij}}{\pi_{ij} M_i} \right\} E(I_{ij}) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left\{ \frac{Y_{ij}}{\pi_{ij} M_i} \right\} \pi_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left\{ \frac{Y_{ij}}{M_i} \right\}$$

Noting that $Y_{ij} = Y_i$ for all M_i households in the i-th SBN

$$E(\hat{t}_{HH}) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \left\{ \frac{Y_{ij}}{M_i} \right\} = \sum_{j=1}^{N} \left\{ \frac{M_i Y_i}{M_i} \right\} = \sum_{i=1}^{N} Y_i = t_{HH}.$$

The variance of $\hat{t}_{HH}$ can be obtained using the standard formula

$$\sum_{i=1}^{N} \sum_{j=1}^{M_i} \sum_{i'=1}^{N} \sum_{j'=1}^{M_i} \frac{\pi_{ij} \pi_{i'j'} - \pi_{iji'j'}}{\pi_{ij} \pi_{i'j'}} \frac{Y_{ij}}{M_i} \frac{Y_{i'j'}}{M_{i'}},$$

where $\pi_{iji'j'}$ is the second inclusion probability of the j-th household in the i-th block and the j-th household in the i-th block. As compared to the variance of the Horvitz-Thompson's estimator $\hat{t}_{HT} = \sum_{i}^{n} \frac{y_i}{\pi_i} = \sum_{i=1}^{N} \sum_{j=1}^{M_i} \frac{I_{ij}}{\pi_{ij}} H_{ij}^{(Res)}$, whose variance is $\sum_{i=1}^{N} \sum_{j=1}^{M_i} \sum_{i'=1}^{N} \sum_{j'=1}^{M_i} \frac{\pi_{ij} \pi_{i'j'} - \pi_{iji'j'}}{\pi_{ij} \pi_{i'j'}} H_{ij}^{(Res)} H_{i'j'}^{(Res)}$, we can see that the only difference is

that we replace $H_{ij}^{(Res)}$ by $\frac{Y_{ij}}{M_i}$. Since the latter tends to be less variable for the households in the same block (ideally, zero variability) as compared to the homeless count from each individual household, we expect that in our stratified simple random sampling design, the multiplicity estimator should have a smaller variance as compared to the Horvitz-Thompson estimator.

2.8. Producing the Estimates

Hidden homeless estimates were produced using SUDAAN (Version 10), a statistical software package developed by RTI International that specializes in providing efficient and accurate analysis of data from complex studies. The estimated total number of hidden homeless persons in the Los Angeles CoC was produced using the DESCRIPT procedure in SUDAAN. A stratified with replacement (STRWR) design nested by stratum was specified in the procedure. A weight statement was also used in the procedure to account for varying selection probabilities in the telephone sample. The final calibrated household sample weight (without adjustment for neighborhood reporting) was used to produce the total number of hidden homeless individuals in a fashion similar to the way it was obtained for HC09. To obtain the total number of hidden homeless individuals using the multiplicity-based approached, a sample weight was used that applied an adjustment for the size of each respondent's SBN – specifically, the conventional calibrated HH weight for each respondent was divided by the respondent's estimated SBN count of HHs (or its multiplicity). Taylor series linearization methods are employed for robust variance estimation of descriptive statistics in the DESCRIPT procedure.

3. Results

The estimated total of hidden homeless persons in the Los Angeles CoC using the Horvitz-Thompson method in 2011 was 10,800 (SE = 3,421) and was based on the entire sample of 3,390 completed interviews in which only 13 households responded that a hidden homeless person was present (see Table 3). This estimate was comparable to the HC09 estimate of 9,451 (SE = 2,339). Again, a household could only qualify as having a hidden homeless person if they had someone living on their property in an unconverted garage, a back porch, or in an encampment, camper or car. These individuals were considered homeless by HUD and estimated counts were added to the total homeless estimate.

The hidden homeless estimate based on the multiplicity-based approach (see Table 4) led to a substantial increase in the hit rate of hidden homelessness ($n = 322$), which

Table 3. *Hidden homeless estimate using conventional (Horvitz-Thompson) method*

Survey year	Overall Hidden Homeless Estimate (Personal Residence)			
	Raw count	Weighted estimate	Weighted standard error	Relative standard error
2011	13	10,800	3,421	32%
2009	16	9,451	2,339	25%

Table 4.   Hidden homeless estimate using multiplicity-based method

Survey year	Overall Hidden Homeless Estimate (Neighborhood Network)			
	Raw count	Weighted estimate	Weighted standard error	Relative standard error
2011	322	18,622	2,889	15%

improved the precision around the weighted estimate as demonstrated by a major reduction in the relative standard error (32% vs. 15%). The estimated total, however, was substantially higher (10,800 vs. 18,622) but was not significant based on a normal distribution test ($p = 0.06$).

3.1.   Quality of Reported Data

The success of the neighborhood reporting approach, however, hinges on the accuracy of the data reported by respondents. Respondents were asked to identify homeless persons living on their personal property and persons living on their neighbor's property (immediate neighborhood). When asked about homeless persons living on a neighbor's property, respondents were also required to remark whether they were *very sure*, *somewhat sure*, or *not very sure at all*. To ensure that the count was computed using quality responses, only hidden homeless individuals reported by respondents who were *very sure* and *somewhat sure* of the status and number of homeless individuals residing on their neighbors' properties were counted. Of the 118 households who provided a response to the question and reported homeless persons living on their neighbor's property, 101 (86%) respondents were *very sure* of the status and number of homeless individuals residing in their immediate neighborhood, while just 17 (14%) were *somewhat sure*. Confidence in the quality of the counts reported by respondents is gained based on this knowledge. An inconsistency may however be present in the respondents' reporting of the size of their street block neighborhood (SBN).

3.2.   Simulations

We have conducted extensive simulation studies to compare the performance of the multiplicity-based estimator versus the Horvitz-Thompson estimator. To imitate the actual LA homeless study, we generated a population of N households randomly assigned to d blocks. A proportion π of these households were assumed to have hidden homeless people on property. We used simple random sampling to select n of these households to collect whether there was a homeless person in the selected property. For comparison, we estimated the total count of homeless using the Horvitz-Thompson estimator and the multiplicity estimator. In the simulation studies, we varied the number of blocks (N), the hidden homeless proportion (π) and the selected household number (n). The results from 1,000 replicates are summarized in the table in the Appendix.

It is clear from the table that both the Horvitz-Thompson estimator and the multiplicity-based estimator produced a precise estimate of the total count of the homeless. However, the multiplicity-based estimator has a greater efficiency gain: The ratio of the mean

squared errors of the Horvitz-Thompson estimator versus the multiplicity-based estimator ranges from 15 to about 20, while the ratio of the relative standard error is between 4 and 5. The efficiency gain of the multiplicity-based estimator seems to increase with increasing sample size and the prevalence of the homeless. All simulation work was done in SAS (Version 9.3).

4.   Discussion

Our estimation approach produced hidden homeless estimates within the methodological framework utilized in the 2009 Los Angeles Homeless Count (HC09) and successfully employed an alternative, multiplicity-based method in the 2011 Los Angeles Homeless Count (HC11). The Horvitz-Thompson (HT) estimate in HC09 (9,451 hidden homeless) is comparable to the HT estimate in HC11 (10,800 hidden homeless). As similar as these two estimates are, they have large standard errors thereby producing wide intervals at the 95 percent confidence level (HC09 ± 4,584; HC11 ± 6705). In terms of relative standard errors, both estimates are not very reliable (HC09 = 25%; HC11 = 32%), thus providing the motivation for multiplicity sampling.

   Our approach takes the simplest form of multiplicity sampling in that we only have to adjust for the size of the network, which here was operationalized as a respondent's street-block neighborhood. Unlike other network approaches (viz., snowball or respondent-driven sampling), we did not have to interview the subjects of our investigation (i.e., the hidden homeless), only count them. To produce unbiased estimates using a multiplicity-based approach, eligible respondents need only to report the size of their network (n) which is then weighted by the reciprocal $1/n$. The main benefit of such an approach is the reduction of sampling error due to the increase in sample size. When costs are fixed, the multiplicity-based approach, we argue, will produce a more precise estimate of a rare event, as we found in the present study (RSE 32% in HT versus RSE 15% in MB). While significantly decreasing the amount of variance around our estimate, we can also see that the multiplicity approach produced a more sensitive measure that detected an additional 7,822 cases of hidden homeless (versus 10,800). Consequently, the multiplicity-based approach produced a more sensitive estimate of the hidden homeless population in Los Angeles, which was more precise than the conventional approach and should be implemented in the next homeless enumeration. Sudman and Freeman (1988), however, suggest that network sampling will lose it attractiveness as the proportion of an event in the population grows and that more conventional methods (such the Horvitz-Thompson estimator) will remain superior. Future research should explore what that cut-off in the population is likely to be and under what conditions it best applies.

   Future work should also look for ways to improve the measure of network size needed to adjust for multiplicity. Recall that in our case, to adjust for the wider reporting framework developed through the multiplicity approach all sample weights were reduced by a factor of the size of each respondent's SBN. Initial framing of the neighborhood network size question required respondents to provide an estimate of how many houses/units were on their block or in their development. Because the selection of random telephone numbers allowed the chance of a house, apartment, or mobile home to be included in the sample, neighborhood network size questions were tailored to household

type. Respondents living in homes or townhomes were asked to provide the number of homes on their block, which consisted of both sides of a street between the two nearest intersections. Respondents living in condominiums or an apartment complex were asked to approximate how many units were in their building and how many buildings were in their complex. The purpose of these questions was to quantify the size of each SBN for which the respondents were reporting hidden homeless. However, the analysis of the data showed that the framing of the question may have been confusing to the complex respondents. In fact, some of these respondents were initially unable to estimate the size of their SBN. As a consequence, imputation techniques and callbacks were used to retrieve missing and erroneous neighborhood size values for nearly a quarter of the households reporting hidden homeless persons. Callbacks were issued for cases where abnormally high values of neighborhood network size were observed. Distinguishing between which measures were valid and which were erroneous, an inherently subjective task, required all neighborhood size data to be re-evaluated. In general, any respondent reporting their SBN size as having more than 500 households/housing units was flagged for further investigation. Nearly all the cases where a respondent reported their SBN size as being more than 500 households/housing units also identified their housing type as a condo or apartment. A potential limitation to this approach might be the difficulties some respondents have estimating SBN. Further research should examine ways to reduce this error, perhaps by shrinking large multi-unit complexes into smaller, manageable units, such as the respondents' street block as contained within the complex, thus reducing the area to be covered within large multiunit complexes. Successfully addressing these limitations holds the promise of making multiplicity-based estimation a more stable, statistically precise and preferred approach for estimating hidden homeless individuals in future homeless counts.

Appendix

Appendix. *Summary of Simulation Study Results for Comparing the Horvitz-Thompson Estimator and Multiplicity-based Estimator*

Total Number of Blocks (d)	Total Household Population (N)	Proportion of Households with Hidden Homeless (pi)	Sample Size (n)	Horvitz Thompson Estimation				Multiplicity Estimation				Ratio of MSE's (HT/Multiplicity)	Ratio of RSE's (HT/Multiplicity)
				Mean Estimate	Standard Error (SE)	Mean Squared Error (MSE)	RSE	Mean Estimate	Standard Error (SE)	Mean Squared Error (MSE)	RSE		
1407	28147	0.24%	15	63.80	63.80	115,665.54	533.29%	68.73	56.00	7,665.01	127.44%	15.09	4.18
			30	73.18	71.36	70,371.81	362.61%	81.18	53.74	3,707.87	74.94%	18.98	4.84
			60	60.99	58.97	28,435.95	276.53%	57.21	33.71	1,372.58	64.75%	20.72	4.27
		1.00%	15	277.72	265.53	521,698.41	260.15%	289.68	153.51	29,173.82	58.98%	17.88	4.41
			30	247.69	228.80	229,999.74	193.35%	279.14	108.06	12,270.08	39.69%	18.74	4.87
			60	288.98	246.82	130,401.42	125.02%	286.98	79.87	6,332.24	27.73%	20.59	4.51
		5.00%	15	1,313.53	1,090.37	2,204,319.09	112.90%	1,410.52	346.04	129,631.61	25.52%	17.00	4.42
			30	1,481.47	1,005.06	1,275,880.84	76.13%	1,420.05	251.33	63,005.29	17.67%	20.25	4.31
			60	1,431.74	759.84	603,227.92	54.26%	1,410.05	174.70	32,186.86	12.73%	18.74	4.26
		20.00%	15	5,644.41	2,815.31	8,399,557.01	51.36%	5,696.20	641.42	429,791.04	11.51%	19.54	4.46
			30	5,518.69	2,017.83	3,945,789.56	36.01%	5,548.82	447.20	194,043.87	7.94%	20.33	4.54
			60	5,581.55	1,440.42	2,203,058.07	26.60%	5,565.22	327.95	116,631.98	6.14%	18.89	4.33
14073	281474	0.24%	150	656.77	590.55	1,291,528.07	172.82%	716.97	254.94	67,062.01	36.12%	19.26	4.78
			300	647.39	533.92	614,409.81	121.13%	651.85	171.34	30,214.91	26.68%	20.33	4.54
			600	643.64	461.04	287,098.63	83.20%	668.87	123.39	14,705.35	18.14%	19.52	4.59
		1.00%	150	2,951.72	2,024.15	5,414,402.82	78.69%	2,810.92	504.40	262,338.27	18.22%	20.64	4.32
			300	2,737.80	1,497.02	2,647,774.81	59.45%	2,765.29	358.28	127,539.36	12.92%	20.76	4.60
			600	2,796.44	1,112.99	1,313,745.34	41.01%	2,789.27	254.44	63,046.37	9.00%	20.84	4.55
		5.00%	150	13,681.51	4,867.70	23,068,629.55	35.11%	13,792.43	1,103.65	1,193,045.01	7.92%	19.34	4.43
			300	13,932.02	3,498.65	12,355,038.88	25.24%	13,948.11	789.48	680,593.33	5.91%	18.15	4.27
			600	14,208.34	2,507.32	5,930,299.17	17.15%	14,165.35	558.09	288,850.02	3.79%	20.53	4.52
		20.00%	150	56,820.22	9,207.34	85,458,036.39	16.26%	56,450.85	2,056.43	4,270,158.51	3.66%	20.01	4.44
			300	56,504.03	6,502.76	43,546,298.18	11.68%	56,479.35	1,457.97	2,293,844.94	2.68%	18.98	4.36
			900	56,335.77	3,752.71	14,610,470.53	6.78%	56,435.06	840.18	717,013.87	1.50%	20.38	4.52
140737	2814740	0.24%	1500	6,867.97	3,437.44	12,299,052.30	51.09%	6,818.99	796.89	655,087.14	11.87%	18.77	4.30
			3000	6,817.30	2,481.81	6,100,243.77	36.24%	6,765.28	561.85	324,008.61	8.42%	18.83	4.31
			6000	6,683.13	1,753.55	2,954,527.12	25.70%	6,768.77	396.87	156,173.25	5.84%	18.92	4.40
		1.00%	1500	28,046.07	7,160.85	50,203,412.83	25.27%	28,230.96	1,616.93	2,657,546.94	5.78%	18.89	4.37
			3000	27,917.53	5,070.78	26,880,793.49	18.57%	28,050.11	1,141.13	1,422,574.90	4.25%	18.90	4.37
			6000	28,073.75	3,603.42	13,275,478.46	12.98%	28,000.56	805.17	639,259.30	2.86%	20.77	4.55
		5.00%	1500	141,600.19	15,863.63	244,333,568.30	11.04%	141,821.14	3,551.83	13,458,268.96	2.59%	18.15	4.27
			3000	140,355.13	11,177.59	127,111,006.08	8.04%	140,338.58	2,492.92	6,670,671.46	1.84%	19.06	4.36
			6000	141,754.06	7,943.89	65,035,869.35	5.67%	141,007.95	1,770.73	3,330,679.41	1.29%	19.53	4.38
		20.00%	1500	565,072.19	29,106.27	843,594,702.23	5.13%	563,609.38	6,501.05	41,463,829.65	1.14%	20.35	4.49
			3000	560,960.79	20,526.31	456,460,140.76	3.81%	561,647.27	4,594.29	21,382,841.57	0.82%	21.35	4.62
			6000	564,178.04	14,546.30	231,048,663.33	2.69%	563,461.16	3,250.42	10,061,108.72	0.56%	22.96	4.78

Note: RSE is the relative standard error defined as the ratio between the standard error and the total count estimate.

5.   References

The American Association for Public Opinion Research (2011). Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys, (7th edition): AAPOR.

Birnbaum, Z.W. and Sirken, M.G. (1965). Design of Sample Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates. Vital and Health Statistics, Series 2, No. 11. DHEW publication no (PHS) 1000. Washington, DC: U.S. Government Printing Office, 1–8.

Blaise 4.8 [computer software] (2007). Voorburg/Heerlen: Statistics Netherlands.

Bring Los Angeles Home: The Campaign to End Homelessness (2006). Los Angeles Housing Services Authority, Web. Available at: http://www.bringlahome.org/docs/BRINGLAHOME_book_final.pdf (accessed January 6, 2010).

Flores-Cervantes, I. and Kalton, G. (2008). Methods for Sampling Rare Populations in Telephone Surveys. In Advances in Telephone Survey Methodology, J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. de Leeuw, L. Japec, P.J. Lavrakas, M.W. Link, and R.L. Sangster (eds). New York: Wiley and Sons.

Kalsbeek, W.D. and Agans, R.P. (2008). Sampling and Weighting in Household Telephone Surveys. In Advances in Telephone Survey Methodology, J.M. Lepkowski, C. Tucker, J.M. Brick, E.D. de Leeuw, L. Japec, P.J. Lavrakas, M.W. Link, and R.L. Sangster (eds). New York: Wiley and Sons.

Kalton, G. (2009). Methods for Oversampling Rare Subpopulations in Social Surveys. Survey Methodology, 35, 125–141.

Kish, L. (1965). Survey Sampling. New York: Wiley and Sons.

Link, B.G., Susser, E., Stueve, A., Phelan, J., Moore, R.E., and Struening, E. (1994). Lifetime and Five-Year Prevalence of Homelessness in the United States. American Journal of Public Health, 84, 1907–1912, DOI: http://www.dx.doi.org/10.2105/AJPH.84.12.1907.

Los Angeles Homeless Services Authority (2009). Greater Los Angeles Homeless Count Report. Available at: http://www.lahsa.org/docs/2011-Homeless-Count/HC11-Detailed-Geography-Report-FINAL.PDF (accessed October 13, 2012).

Potter, F. (1988). Survey of Procedures to Control Extreme Sampling Weights. Proceedings of the American Statistical Association Section on Survey Research Methods, 453–458.

SAS Statistical Software [computer program]. (Version 9.3), Cary, North Carolina.

SUDAAN [computer program] (Version 10). Research Triangle Park, North Carolina.

Sudman, S. and Freeman, H.E. (1988). The Use of Network Sampling for Locating the Seriously Ill. Medical Care, 26, 992–999.

Tompsett, C.J., Toro, P.A., Guzicki, M., Schlienz, N., Blume, M., and Lombardo, S. (2003). Homelessness in the US and Germany: A Cross-National Analysis. Journal of Community and Applied Social Psychology, 13, 240–257, DOI: http://www.dx.doi.org/10.1002/casp.724.

Tompsett, C.J. and Toro, P.A. (2004). Public Opinion. Encyclopedia of Homelessness: SAGE Publications. Available at: http://sageereference.com/homelessness/Article_n133.html (accessed January 18, 2010).

Toro, P.A. and Janisse, H.C. (2004). Homelessness, patterns of. In D. Levinson (ed.). Encyclopedia of homelessness (pp 244–250) Thousand Oaks, CA: Sage.

Toro, P.A., Tompsett, C.J., Lombardo, S., Phillippot, P., Nachtergael, H., Galand, B., Schlienz, N., Stammel, N., Yabar, Y., Blume, M., MacKay, L., and Harvey, K. (2007). Homelessness in Europe and the United States: A Comparison of Prevalence and Public Opinion. Journal of Social Issues, 63, 505–524, DOI: http://www.dx.doi.org/10.1111/j.1540-4560.2007.00521.x.

Toro, P.A. (2005). Community Psychology: Where Do We Go from Here? American Journal of Community Psychology, 35, 9–16, DOI: http://www.dx.doi.org/10.1007/s10464-005-1883-y.

Toro, P.A. (2006). Trials, Tribulations, and Occasional Jubilations While Conducting Research With Homeless Children, Youth, and Families. Merrill-Palmer Quarterly, 52, Academic OneFile. Available at: http://find.galegroup.com/gtx/start.do?prodId=AONE&userGroupName=unc_main. (accessed January 11, 2010).