The Design of Homeless Surveys

Ronaldo Iachan and Michael L. Dennis, Research Triangle Institute
Ronaldo Iachan, PO Box 12194, Research Triangle Park, NC 27709-2194

KEY WORDS: rare mobile population, street sample, shelter sample, temporal stratification

1. Introduction

Over the last decade, growing attention has been devoted to the homeless problem which has been steadily increasing in scale as it has changed in character. Several studies have investigated the numbers and characteristics of homeless individuals in different areas of the country.

This paper explores design issues relevant to homeless surveys. These issues are discussed and illustrated with a study of homeless people in the Washington, DC, metropolitan area currently conducted by the Research Triangle Institute as part of the National Institute on Drug Abuse's (NIDA, 1989) Washington, DC, Metropolitan Area Drug Survey (DC*MADS) study. The DC-homeless survey includes shelter, street and service location components.

Two issues that should be considered in the design of surveys of homeless street people pertain to the need for stratification both by geographic location and by time. Spatial stratification is necessary to help locate eligible members of this population, which is both rare and mobile. Temporal stratification allows both the computation of estimates of prevalence and incidence, and of seasonal variations and trends. A third issue relates to the use of multiple sampling frames, to capture populations components in such disparate settings as shelters, service locations, and streets.

Shelter surveys only capture a small portion of the homeless population and do not properly represent subgroups of potential interest (Dennis & Iachan, 1991). For example, Davidson (1991) found that the rates of substance abuse, mental illness and mental retardation among 313 people served by nine shelters on one day in July were significantly different from the rates among the 632 who were on the shelters "do not admit" list on the same day. Several studies have attempted to address this bias by supplementing shelter surveys with samples of people drawn from other locations. Dennis (1991) categorized 14 homeless studies into three camps. Studies in the first camp use only samples of service system locations (e.g., shelters, soup kitchens, day programs) because they are cheaper and cover most of the population (e.g., Breakey et al. 1989; Burt & Cohen, 1989). Those in the second camp consider probability samples of shelter and street locations to reduce the potential for bias due to undercoverage and limitations of service systems (e.g., Rossi et al., 1986). Studies following the third, compromise approach, focus on service system samples but also include either purposive or partial samples of high-density street locations (e.g., Vernez, et al., 1988; Farr et al., 1986; Ringwalt and Iachan, 1990). Only one study, the DC*MADS Homeless Study reviewed in the next section, attempts a comparison of all three approaches.

2. An illustrative homeless survey.

For almost two decades, NIDA has relied on a series of household and hospital surveys to monitor substance abuse in America. While this strategy has been useful as a general barometer of drug use, concern has increased that it underrepresents several subpopulations that are more likely to be adversely affected by substance abuse such as school dropouts, adult and juvenile criminal offenders, the institutionalized, drug abuse treatment clients, pregnant drug abusers and, most notably, the homeless population.

NIDA has contracted with the Research Triangle Institute and three other firms to conduct a series of 16 comprehensive studies under the umbrella of a single research study program, DC*MADS. This effort is an attempt to collect data about drug abuse from all of these subpopulations and the household population during the same year in one metropolitan area. To the extent that it is successful, DC*MADS will be used as a model to collect similar data in other metropolitan areas.

The DC*MADS homeless study component will examine the prevalence, and consequences of drug use in the homeless population. The study will also compare several definitions of homelessness. At one extreme, it includes all people encountered in encampments, shelters and service locations, like the homeless enumeration study conducted by the U.S. Census on March 20, 1990. At the other extreme, the target population can be confined to literally homeless persons who at a given night may be found either in a shelter or on the streets. It may be worth pointing out that neither definition necessarily includes all doubled-up persons in households, institutionalized persons or persons considered at-risk for homelessness.

Using a broader definition, or one that looks at a period of time instead of a single night, can dramatically increase the estimated population size and characteristics. Santiago and colleagues (1988), for instance, found that changing their definition from "currently homeless" to "homeless in the last three months" increased from 106 to 159 (50 percent) the number of people identified as homeless in a sample of psychiatric hospital patients. The NIDA study is examining how sensitive the estimates are to different definitions. Our broadest definition of homelessness includes many people who are precariously housed or living in nontraditional dwellings. In addition to their current episode of homelessness, respondents will be asked about their 12-month prevalence of homelessness.

3. Sample Design

The design of a sample that covers both the shelters and street would ensure complete coverage of the "literally homeless" population, provided that we can design a street sample that gives every street homeless person a non-zero and known probability of selection. The street sampling frame is the negative image of the usual area household sampling frame; however, instead of dwellings its units constitute non-dwelling-units. To the extent that we are logistically unable to locate people in the sampled blocks, however, it may be useful to sample from other locations through which street people are likely to pass (e.g., soup kitchens, jails).

The original sampling design called for independent seasonal shelter samples and street samples in the winter and spring of 1991. After higher than expected risks, lower than expected yields, and higher than expected overlap between the sampling frames, the data collection was extended into the summer (NIDA, 1991). The summer sample supplement included a fourth monthly shelter sample, a street sample from major encampments of homeless people, and a soup kitchen sample. Thus, the final design included: (1) four monthly samples of individuals who spent the night in emergency shelters or hotels for homeless people, (2) three monthly probability samples of people in nondomiciles (i.e., those living in the streets) between 4:00am and 5:30am, (3) one sample of people in encampments, and (4) one sample of people receiving food from the area's soup kitchens, mobile meal programs, and food banks. Table 1 summarizes the study's sampling plan for these components.

The shelter and soup kitchen frames for the DC*MADS homeless study was based on lists maintained by the DC Council of Governments and the Interfaith Council. These lists were verified by phone and supplemented with lists supplied by each local municipality. Sample shelters were selected with probabilities proportional to bed capacity. Soup kitchens were sampled with probabilities proportional

to the number of meals served. An approximately constant number of shelter or soup kitchens clients were selected in each sampled facility. Both frames were stratified by size.

The encampments sample was based on locations where local experts said 5 or more homeless people could be found every night. All locations were verified visually by drivebys on two separate nights. A simple random sample of encampments was then selected.

The samples allow estimation of seasonal trends for the winter and spring, each with two independent samples, and calculation of the ratio of street to shelter people. The sampling design includes both temporal and spatial dimensions.

4. Street Sampling

Ideally a street sample should identify most of the non-domiciled people; however,it is difficult even for knowledgeable people to predict where homeless street people will be sleeping on a given night. The street population is rare, mobile, and elusive. It is difficult to locate people who are actively hiding to avoid both vicitimization and being run off by authorities. Unfortunately provider and advocate estimates of the number of street people are also unreliable and vary by ten-fold or more (Farr et al., 1986; Rossi, 1989; Vernez et al., 1988).

The NIDA street survey was based on a two-stage sample of census blocks. Sample blocks were selected in two stages; first-stage units were census tracts. At both stages, the sample was stratified by the likelihood of finding a homeless person in the area during the predawn or early morning hours. These hours were chosen to minimize the amount of screening needed to identify eligible individuals and to select a time when they would be least mobile. Stratification information sources included local service providers and homeless people.

The first-stage sampling frame was stratified into three categories according to the likely concentration of homeless individuals. Table 2 presents the first-stage stratification and sample allocation. The second-stage, block frame was similarly stratified, and sample blocks were selected with equal probabilities within each stratum. All homeless individuals identified in a sample block during the data collection period were counted and interviewed.

5. Sampling Over Time and Double Counting Issues

Time-related problems that must be addressed in sampling homeless people include:

- seasonal changes, e.g., due to weather,
- changes in the service systems,
- population movement across sampling frames, and
- problems associated with using a currently-homeless definition.

Seasonality affects the number and distribution of homeless people in many ways. In most areas, winter means higher utility bills that force some people out of their homes. On the other hand, warmer spring weather makes sleeping outside a more viable option. The seasonality of the data collection period underlies the relative allocation of the total samples to shelter and street sites. Because more people seek shelter in cold weather, more observations are required from shelters than from the streets in the winter, and conversely in the spring.

We selected independent seasonal samples stratified by month, and randomly assigned shelters and blocks to the sampled nights. The selection of monthly samples prevents clusters of days at the beginning or end of the season. It also minimizes the chance of visiting all of the selected shelters or blocks in one municipality in a same month. More importantly, it permits the computation of monthly and seasonal estimates and trends (Iachan, 1989).

Most of the service systems in the DC area change their level of services around April 1st of each year. The two independent seasonal samples are designed to capture the April 1st change in the service systems. To avoid having too many days clustered around the beginning or end of the month, when entitlement checks and paychecks often arrive in the mail, the temporal sample is stratified by week. Potential biases are further reduced by randomly assigning shelters and blocks to the sampled nights.

In many DC shelters, people are entering shelter buildings from 6:00 p.m. until 6:00 a.m. However, even before everyone is in for the night, many start leaving (up to 50% of shelter clients may have left by 4:00 a.m. to start walking over to a soup kitchen). Thus, there is no one single time in which the entire shelter population for a given night can be captured. It also means that on a given night, the same person may be in different shelter and street frame units. The NIDA survey is addressing the first problem by taking a systematic sample of people as they enter the shelters throughout the night. The same sample nights were used for the street and shelter samples to minimize the overlap between the two components. The chances are negligible that a person can be found in a shelter and then in the street between 4 and 5 AM that same night. The street data collection takes place in a period of relatively low mobility (4:00 to 5:30 a.m.). Finally, the respondents will be asked whether they have ever been interviewed before. The overlap questions will look at where the respondent was during the sampled night, the last 12 months, and over a lifetime.

Determining the overlap between multiple frames is a common problem in designing a sample. When an overlap cannot be defined away, it is necessary to measure it in developing a population estimate. In homeless studies, this has been done by asking people about their sleeping quarters and/or service utilization in the last 7-30 days (e.g., Burt & Cohen, 1989; Farr et al., 1986). A common but more dubious practice is to inflate or extrapolate this number to the last 12 months or a lifetime. The problem with the latter technique is that the same individual often becomes homeless at several points in a year. These episodically homeless people bias the resulting adjustments and produce annual estimates of unique episodes, not unique individuals.

The NIDA study will address these problems by asking respondents about their lifetime, 30-day, and 24-hour utilization of shelters and services (e.g., soup kitchens, clinics) and sleeping on the street. By comparing the estimates for the three different units of time, we can examine the sensitivity of the statistical models that are being used to extrapolate annual estimates.

6. Discussion

The design of the DC*MADS homeless study street component incorporated the knowledge gained during Rossi's (1989) Chicago study and the Census enumeration (S-night). Still, the state-of-the-art design presented several problems and opportunities for further design improvements.

For the design of nighttime surveys, it may be helpful to define two primary subgroups of the street homeless population. The first group consists of those individuals who may be found clustered in encampments and who often tend to seek safety in numbers. The second group includes isolated individuals who are either wandering in drug- or mental-illness- induced stupor or are hiding for safety or privacy reasons. Locating, listing, and sampling the second group is much more difficult and expensive than the first group.

The DC*MADS survey was designed to capture individuals in both of these groups but was only partially successful in covering (or uncovering) members of the second group (NIDA, 1991). This partial coverage occurred despite intensive efforts ranging from going into places of difficult or dangerous access (e.g., abandoned buildings and crack houses) to screening and interviewing in the hours of presumed lowest mobility. In fact, the interviewers were instructed to wait for any person found sleeping in the street to wake up. Nevertheless, a majority of the street people screened and/or interviewed were found in movement. Another finding of relevance is that a great majority of the eligible persons interviewed were regular service users (e.g.,

soup kitchens), a finding that reinforces the notion of sampling daytime service locations.

The service location sampling approach was included in the DC*MADS sample for June 1991. For this purpose, we constructed a comprehensive frame of service programs further subdivided by sites and meals. Sampling units were meal-sites (e.g., breakfast at a particular site). This study component will adopt less strict eligibility criteria (i.e., broader definitions for the different degrees of homelessness), and will throw further light on the overlap between the various homeless subpopulations.

Based on this review and our experience in DC*MADS, there are several alternative street sampling designs that hold some promise for further addressing the cost and/or precision issues related to homeless population surveys. Table 3 compares five potential strategies for sampling street homeless individuals. The relative advantages of each strategy are presented along two basic dimensions: cost and coverage. Other factors to consider include whether the strategy yields a probability sample of areas and homeless people in these areas.

The sampling frame for the first strategy consists of a list of known clusters of street persons that may be verified by field staff. For the second strategy, the frame is restricted to areas with high density ratings provided by expert judgment. To the extent that such judgments are consider sufficient to exclude an area, it will produce a partial probability sample (i.e., a probability sample of the targeted areas). Where it has been used (e.g., Vernez et al., 1988), such expert judgments have been typically verified through drive-by or "windshield" observations.

The third strategy calls for a stratified sample of areas that are then listed to exclude areas unlikely to contain homeless people. Such a procedure is analogous to that used in household surveys and would thus incorporate both expert judgment and direct observation. The fourth option is an adaptive cluster sampling method analogous to the Waksberg-Mitofsky variation of random digit dialing (RDD). The idea, as in the RDD variant, is to reduce the number of screenings needed to find eligible population members. This reduction is achieved by following up on successful screenings in a given cluster. The fifth option is a one or two-stage stratified random sample that incorporates expert information on the probability of identifying homeless people. Of the five listed strategies, only the last three (which are also the most expensive) assign non-zero probabilities of selection to each area. Of these, only the last one has been fully implemented.

REFERENCES

Breakey, W. R., Fischer, P. J., Kramer, M., Nestudt, G., Romanoski, A. J., Ross, A., Royall, R. M., & Stine, O. C. (1989), "Health and Mental Health Problems of Homeless Men and Women in Baltimore." Journal of the American Medical Association, 262(10), 1352-1357.

Burt, M.R., & Cohen, B.E. (1989). America's Homeless: Numbers, Characteristics, and Programs that Serve Them (Urban Institute Report 89-3). Washington, DC: Urban Institute, 2100 M. Street, NW.

Davidson, D. (1991). A Snap Shot Survey of Hard to Serve Homeless Clients in Northern Virginia. Fairfax, VA: Northern Virginia Coalition for the Homeless.

Dennis, M. L. (1991). "Changing the Conventional Rules: Surveying Homeless People in Nonconventional Locations." Housing Policy Debate, 2(3), 1-32.

Dennis, M. L. & R. Iachan (1991). "Sampling Issues in Estimating the Extent of Alcohol, Drug Abuse and Mental Illness Problems," in C. M. Taeuber (Ed.), Enumerating Homeless Persons: Methods and Data Needs, Washington, D.C.: Bureau of the Census, U.S. Department of Commerce, pp. 188-91.

Dennis, M. L., R. Iachan, J. S. Thornberry, & R. M. Bray (1991). "The RTI Method: Sampling Over Time," in C. M. Taeuber (Ed.), Enumerating Homeless Persons: Methods and Data Needs, Washington, D.C.: Bureau of the Census, U.S. Department of Commerce, pp. 167-170.

Farr, R. K., Kogel, P., & Burnam, A. (1986). A Study of Homelessness and Mental Illness in the Skid Row Area of Los Angeles (NIMH Grant No. iHR MH36809-01). Los Angeles, CA: Department of Mental Health.

Iachan, R. (1989). "Sampling in Time and Space." Proceedings of the Section of Survey Research Methods of the American Statistical Association, 636-640.

Hamilton, Rabinovitz & Alschuler, Inc. (1987). The Changing Face of Misery: Los Angeles' Skid Row Area in Transition, (Four volume report to the Community Redevelopment Agency of Los Angeles). Los Angeles, CA: Hamilton, Rabinovitz & Alschuler, Inc.

National Institute on Drug Abuse (1989). The Washington, DC, Metropolitan Area Drug Study, Contract No. 271-89-8340. Rockville, MD. [for further information on this study, contact the NIDA project officer, Ms. Elizabeth Y. Lambert, at 301-443-6667 or the Research Triangle Institute (RTI) project director of DC*MADS, Dr. Robert M. Bray, at 919-541-6433, or the RTI homeless and transient population study leader, Dr. Michael L. Dennis, at 919-541-7136].

National Institute on Drug Abuse (1991). Homeless and Transient Population Study Training Manual. (Technical document under NIDA contract #271-89-8340 "Washington DC Metropolitan Area Drug Study".) Research Triangle Park, NC: Research Triangle Institute.

Ringwalt, C., & Iachan, R. (1990). Design Summary of Proposed Study to Estimate the Characteristics of Runaway and Homeless Youths. Research Triangle Park, NC: Research Triangle Institute.

Rossi, P.H. (1989). Down and Out in America: The Origins of Homelessness. Chicago, IL: The University of Chicago Press.

Rossi, P. H., Fisher, G. A., & Willis, G. (1986). The Condition of the Homeless of Chicago. Amherst, MA, and Chicago, IL: Social and Demographic Research Institute and NORC.

Santiago, J. M. Bachrach, L. L., Berren, M. R., & Hannah, M. T. (1988). Defining the Homeless Mentally Ill: A Methodological Note." Hospital and Community Psychiatry, 39(10):1,100-1,101.

Vernez, G., Burnam, M.A., McGlynn, Trude, S., & Mittman, B.S. (1988). Review of California's Program for the Homeless Mentally Disabled, (Final report under contract no. 86-77219 to the California Department of Mental Health). Santa Monica, CA: Rand Corporation (R-3631-CDMH).

Table 1. Sampling Design Summary for the Street and Shelter Survey Components

Sampling Stage/Unit	Sample Method	Size (Rate)
1. Shelter Sample		
1a. Days	Stratified random sampling	64 days (4 per week)
1b. Shelters	Probability proportional to size (bed capacity)	94 shelters (1-2 per day)
1c. Shelter clients	Systematic random sampling	484 interviews (5-6 per shelter)
2. Street Sample		
2a. Days	Stratified random sampling	48 days (4 per week)
2b. Blocks	Stratified random sampling	432 blocks (9 per day)
2c. Street homeless persons	All eligible individuals found in sample blocks/days	54 interviews (.125 per block)
3. Encampment Sample		
3a. Days	Stratified random sampling	16 days (4 per week)
3b. Encampments	Simple random sampling	16 encampments (1 per day)
3c. Encampment homeless persons	All eligible individuals found in sample encomponent/days	146 interviews (9-10 per encampment)
4. Soup Kitchen Sample		
4a. Days	Stratified random sampling	16 days (4 per week)
4b. Soup Kitchens	Probability proportional to size (meal capacity)	32 kitchens (2 per day)
4c. Soup Kitchen clients	Systematic random sampling	200 interviews (6-7 per kitchen)

Source: Adapted from NIDA, 1991

Table 2. First-stage Stratification and Sample Allocation for Street Sample

Municipality	Tract Ratings: High	Medium	Low	Total
(a) Population counts:				
Alexandria	--	3	30	33
Arlington	3	2	34	39
Charles	--	4	10	14
Calvert	--	3	7	10
Fairfax City	--	--	5	5
Fairfax Co.	11	--	131	142
Frederick	3	11	18	32
Montgomery	12	11	126	149
Prince George's	51	60	61	172
Prince William	7	5	19	31
Manassas City	--	2	2	4
Manassas Park	--	2	--	2
Falls Church	--	2	1	3
Loundon	4	3	9	16
DC	15	9	159	183
Stafford	--	--	5	5
	106	117	617	840
(b) Sample Sizes				
Alexandria	--	1	--	1
Arlington	--	--	1	1
Charles	1	1	--	2
Dist. of Columbia	3	4	6	13
Fairfax	5	--	5	10
Frederick	3	2	1	6
Manassas Park	--	1	--	1
Montgomery Co.	3	--	2	5
Prince George's	15	6	1	22
Prince William	2	1	--	3
	32	16	16	64

Source: (Dennis, Iachan, Thornberry & Bray, 1991).

Table 3. Potential Strategies for Street Sampling

Strategy	Cost	Coverage	Geographic Probability Sample	Applications
Listing/sampling encampments	Low	No loners	No	Rossi's supplement DC*MADS supplement
Sampling high-density areas	Moderate	No low-blocks	Partial	Vernez et al. (1988)
Sampling and listing	High	Very Good	Yes	---
Geographic analog of Waksberg-Mitofsky	High	Fair	Yes	---
Stratified Random Sample	High	Good	Yes	Rossi et al (1986) NIDA (1991) Hamilton et al (1986)