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 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 probability 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 15 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 probability of selection. Instead of dwellings, 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 housing, 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 victimization 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
capturing 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 home-
less 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 more viable option. The seasonality of the
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 6:30 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 produc
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 utili-
ization of shelters and services (e.g., soup kitchens, clinics)
and sleeping on the street. By comparing the estimates for the
different frames of time, we can examine the sensitiv-
ity of the statistical models that are being used to extrapolate
annual estimates.

6. Discussion

The design of the DC*MADS homeless study street
clean 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 development.

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

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 considered 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 "windshielded" 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-Milofsky 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 the 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


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 301-443-8667 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].


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

<table>
<thead>
<tr>
<th>Sampling Stage/Unit</th>
<th>Sample Method</th>
<th>Size (Rate)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Shelter Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1a. Days</td>
<td>Stratified random sampling</td>
<td>64 days (4 per week)</td>
</tr>
<tr>
<td>1b. Shelters</td>
<td>Probability proportional to size (bed capacity)</td>
<td>94 shelters (1-2 per day)</td>
</tr>
<tr>
<td>1c. Shelter clients</td>
<td>Systematic random sampling</td>
<td>484 interviews (5-6 per shelter)</td>
</tr>
<tr>
<td>2. Street Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2a. Days</td>
<td>Stratified random sampling</td>
<td>49 days (4 per week)</td>
</tr>
<tr>
<td>2b. Blocks</td>
<td>Stratified random sampling</td>
<td>432 blocks (9 per day)</td>
</tr>
<tr>
<td>2c. Street homeless persons</td>
<td>All eligible individuals found in sample blocks/days</td>
<td>54 interviews (.126 per block)</td>
</tr>
<tr>
<td>3. Encampment Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3a. Days</td>
<td>Stratified random sampling</td>
<td>16 days (4 per week)</td>
</tr>
<tr>
<td>3b. Encampments</td>
<td>Simple random sampling</td>
<td>18 encampments (1 per day)</td>
</tr>
<tr>
<td>3c. Encampment homeless persons</td>
<td>All eligible individuals found in sample encampment/days</td>
<td>146 interviews (9-10 per encampment)</td>
</tr>
<tr>
<td>4. Soup Kitchen Sample</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4a. Days</td>
<td>Stratified random sampling</td>
<td>16 days (4 per week)</td>
</tr>
<tr>
<td>4b. Soup Kitchens</td>
<td>Probability proportional to size (meal capacity)</td>
<td>32 kitchens (2 per day)</td>
</tr>
<tr>
<td>4c. Soup Kitchen clients</td>
<td>Systematic random sampling</td>
<td>200 interviews (5-7 per kitchen)</td>
</tr>
</tbody>
</table>

Source: Adapted from NIDA, 1991
Table 2. First-stage Stratification and Sample Allocation for Street Sample

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Tract Ratings</th>
<th></th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
</tr>
<tr>
<td>Alexandria</td>
<td>--</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>Arlington</td>
<td>3</td>
<td>2</td>
<td>34</td>
</tr>
<tr>
<td>Charles</td>
<td>--</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Calvert</td>
<td>--</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>Fairfax City</td>
<td>--</td>
<td>--</td>
<td>5</td>
</tr>
<tr>
<td>Fairfax Co.</td>
<td>11</td>
<td>--</td>
<td>131</td>
</tr>
<tr>
<td>Frederick</td>
<td>3</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>Montgomery</td>
<td>12</td>
<td>11</td>
<td>128</td>
</tr>
<tr>
<td>Prince George's</td>
<td>51</td>
<td>60</td>
<td>61</td>
</tr>
<tr>
<td>Prince William</td>
<td>7</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>Manassas City</td>
<td>--</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Manassas Park</td>
<td>--</td>
<td>2</td>
<td>--</td>
</tr>
<tr>
<td>Falls Church</td>
<td>--</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Loundon</td>
<td>4</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>DC</td>
<td>15</td>
<td>9</td>
<td>159</td>
</tr>
<tr>
<td>Stafford</td>
<td>--</td>
<td>--</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>117</td>
<td>617</td>
</tr>
</tbody>
</table>

(b) Sample Sizes

<table>
<thead>
<tr>
<th>Municipality</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexandria</td>
<td>--</td>
<td>1</td>
</tr>
<tr>
<td>Arlington</td>
<td>--</td>
<td>1</td>
</tr>
<tr>
<td>Charles</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Dist. of Columbia</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Fairfax</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Frederick</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Manassas Park</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Montgomery Co.</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Prince George's</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>Prince William</td>
<td>2</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>32</td>
<td>18</td>
</tr>
</tbody>
</table>

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

Table 3. Potential Strategies for Street Sampling

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