Sample Design Options for a Survey of Homeless in Canada

Harold Mantel and Wesley Yung

1. Introduction

Statistics Canada (STC) has been approached by Human Resources Development Canada (HRDC) to look into the feasibility of conducting a survey of homelessness in several Canadian cities. The primary objective of this survey would be to obtain accurate counts of the numbers of homeless persons. Initially the survey is planned to get a single count, but the survey may become regular, perhaps being conducted more than once per year to explore seasonal patterns of homelessness.

The population of interest for the survey is those who are absolutely homeless according to the 1987 United Nations definition – *i.e.*, individuals or families who either have no housing at all or are staying in temporary forms of shelter – within certain specified Canadian cities. The component of the homeless population staying temporarily in private residences of friends or acquaintances will not be covered by this survey. On any particular night the population of interest can usefully be divided into two components – the shelter population and the street population. The shelter population means those homeless who are staying in some form of temporary shelter including shelters for the homeless, battered women's shelters, alcohol or drug recovery residences, *etc*. The street population refers to people staying overnight literally on the street, or in abandoned buildings, parks, bus shelters, or other locations not intended for human habitation. The street population can be further subdivided into those who have some contact with services such as soup kitchens and drop-in centres, and those who do not have any such contact. The "service" portion of the street population can potentially be contacted and surveyed through the services that they use. The "non-service" portion of this population will be much more difficult and expensive to survey.

Given that the homeless population can be divided into three sub-populations (shelter, services, and the street population), it is recommended that a multi-frame approach be adopted for the current study. With this approach, we construct sample frames for each of the three components of the homeless population and develop sample designs for each of the components. Since each component will have it's own sampling frame, the sample design for each component can be different depending on the particular characteristics of the sub-population. Unfortunately, the multi-frame approach does create some problems in terms of overlapping units. It is possible, and probably likely, that some individuals will be covered by more than one sampling frame. For example, an individual spending the night in a shelter may have visited a soup kitchen during the day and thus would be covered by the shelter and the services sampling frames. In order to avoid double counting this individual, some method of unduplication needs to be developed. This method could be as simple as asking the respondent if he/she has used a shelter/soup kitchen during the day or as complicated as deriving a unique identifier for each respondent and using this identifier to unduplicate the collected data. Reliably identifying this overlap is likely to be one of the most difficult challenges facing the survey.

In this short note, we present some sample design options for the HRDC/STC study on homelessness. We apply a multi-frame approach and present sample design options for each of the three components. Finally, we include some general issues that need to be addressed before the implementation of any of the sample design options given in this document. This document is a follow-up to two earlier documents that described the objectives of the survey (homeless_draft_objectives4.doc) and a literature review on the topic of surveys of homelessness (lit_rev2.doc).

2. Shelter Component

Target Population: The target population for the shelter component is all homeless people making use of a shelter for the homeless during the reference period of the survey. However, data for a significant portion of this population may become available through the Canadian Mortgage and Housing Corporation's (CMHC) Homeless Individuals and Families Information System (HIFIS). HIFIS was designed to provide administrative data on shelter usage to different levels of government. For those shelters where it is implemented HIFIS will collect a rich variety of variables concerning demographics, health, reasons for homelessness, *etc.*

While it appears that the HIFIS system will be a rich source of data for the HRDC/STC study, there are some concerns about the implementation schedule, the coverage of the system, the variables collected, and the usability of the collected data. Given these concerns, it would be prudent to consider options for a sample survey of shelters. This sample design should take into account the uncertainty surrounding the HIFIS system (i.e. implementation schedule and if implemented, the coverage). With this in mind, we divide the universe of shelters into two populations: 1) the population of shelters covered by HIFIS and 2) the population of shelters to be covered by a sample survey. We will assume that the data on the population of shelters covered by HIFIS will be gathered by the HIFIS system and that these shelters will not be eligible to be sampled. Nevertheless, the sample design should be flexible enough to allow the sampling of HIFIS shelters. Another possibility is that sampling be conducted in the HIFIS shelters, with the HIFIS data being used as auxiliary data to improve estimates.

Thus, in this document we will concentrate on the non-HIFIS population with the knowledge that a sample of HIFIS shelters, if needed, can be obtained using similar techniques. In terms of homeless individuals these two populations may have some overlap since, depending on the reference period, an individual can be in both populations. Thus it will be necessary to identify and account for this overlap in the estimation.

Reference period/collection period: The reference period for the survey will affect the composition of the target population in three ways. It will affect the overall population of homeless individuals, since it will include any person who is homeless at any time during the reference period. Similarly, the component of the overall homeless population that makes use of shelters will also be affected.   Finally, the overlap between the HIFIS and non-HIFIS components of the shelter population may also be affected.

The reference and collection periods also determine the possibilities for an individual to be in the survey more than once, either in the same frame or in different frames. Some information about these multiple possibilities will need to be collected for individuals in

order to derive unbiased estimation weights. More discussion is given below in Section 5 on global issues.

Sampling frame: It is assumed that the shelter population of homeless will be surveyed through the shelters that they use. In order to cover the non-HIFIS population, a list or sampling frame of shelters will need to be created. This sampling frame should include all shelters of interest in the geographical region of interest (city, region, etc…). This should include any hotels where homeless may be staying, providing they can be identified. During creation of the sampling frame, it will be necessary to define the types of shelters (youth shelters, battered women shelters, etc…) that make up the target population and to ensure that **all** shelters of these types are included on the sampling frame. If HIFIS is implemented, then the shelters covered by HIFIS will have to be identified and removed from the non-HIFIS sampling frame.

The information to be included on the sampling frame for each shelter should indicate the type of shelter (*e.g.* youth shelters, battered women shelters), an approximate size measure (*e.g.* number of beds or typical number of clients), address and contact information, and an indicator of whether or not information from the shelter is available through HIFIS.

Sample Design: After creation of the sampling frame, stratification, a commonly used technique to improve the efficiency of sample designs, should be considered. The three major American studies (The Chicago Study, The Urban Institute Study and the DC*MADS Study) all used stratification by size (shelter capacity or the number of beds) in their shelter components, and we propose some similar type of stratification for this survey. We may also wish to stratify by type of shelter. The number of strata will need to be decided upon, as well as the stratum boundaries that should be data driven but can be fixed by the client if desired. If the boundaries are to be data driven, techniques such as the cumulative root-*f* rule or the optimal method of Hidiroglou-Lavallée could be applied.

After stratification, the sample must be allocated to the strata. However, before the allocation can be performed, the final sample size must be determined. The final sample size may depend on operational considerations (cost of data collection, length of collection period, coverage of the HIFIS system, *etc…*) as well as statistical considerations (precision of estimates, response rates, *etc…*). Note that the final sample size should reflect the probability of obtaining a response from the selected individual. Once a sample size is determined, it can be allocated to the strata using optimal or power allocation. It may be desirable to designate certain strata (such as large shelter strata) as take-all strata. That is, all shelters in the take-all strata are to be sampled with certainty.

Once the allocation of the sample to the strata has been performed, a random sample of shelters can be selected from the sampling frame, perhaps using a probability proportional to size (PPS) sampling method. Sub-sampling within selected shelters can be performed in two ways: 1) all clients can be selected from within the sampled shelter or; 2) a sample of clients can be selected from within the sampled shelter. For the current study, a complete enumeration (option 1) could be used in small shelters, while a systematic sample of clients could be drawn in the large shelters. To draw a systematic sample, a list or roster of clients would be needed. Given a roster, a sampling interval could be derived by dividing the number of individuals on the roster by the desired sample size for that shelter. A random start point is then chosen and that individual

3

would be selected. The sampling interval is then applied to identify the next selected individual. This continues until the desired sample size is reached

In summary, a sample of non-HIFIS shelters can be obtained using a stratified two-phase design. The shelters should be stratified by some measure of size (number of beds, capacity or typical usage) and possibly by type of shelter. The first phase sample should consist of a sample of shelters selected with probability proportional to size (again, number of beds, capacity or typical usage). Finally, the second phase sample would consist of a random sample of clients or a full enumeration of all clients in the selected shelters.

3. Services Component

Target population: The second part of the homeless population is the services component, that is, those homeless people who make use of services such as soup kitchens or drop-in centres. This group can be subdivided into those who stayed in one of the shelters covered by the shelter component of the survey, and those who did not. An important question that must be considered is whether those homeless who are in the shelter population should also be considered as part of the target population for the services component. A related question is whether it is permitted for the same individual to be interviewed more than once, possibly for different components of the survey. More discussion of this general issue is given below in Section 5 on global issues.

Reference period/Collection period: The ramifications of the reference and collection periods are similar to what they are for the shelter component, specifically their impact on the populations covered by the different components of the survey and on the possibilities of individuals being selected more than once. These are discussed in more detail in Section 5.

Sampling frame: The services population can be contacted through the services that they use. The sampling unit for a survey of the services population is the recipient of the service itself; for example, a person served a meal in a soup kitchen, or a person found in a drop-in centre at a particular time.

A sampling frame could consist of a complete list of service providers. It may, in fact, be better to consider this as two separate frames – one of soup kitchens and the other containing all other service providers. Note that shelters that also serve meals would need to be included if those meals are served to people who are not staying in the shelter. Additional information, such as estimates of the numbers of clients served, or lists of service periods, could be used to improve the efficiency of sampling. If the service periods are available for each service provider, then for each service period a sample of service providers could be chosen, giving more direct control over the sample size by time period, which could be useful from an operational point of view.

Survey design: A two-phase sampling procedure, similar to the one proposed for the shelter component, is suggested. At the first phase a sample of service providers and service periods would be selected. The information collected at the first phase would be simply the number of clients served during the service period. This first phase could be a two-stage sample, with a second-stage sample of service times selected from a first-stage sample of service providers, or vice-versa, or a more general type of sampling plan could be used. Stratification and/or PPS sampling may also be used. The type of sampling

plans possible for the first phase will depend on the information available on the sampling frame.

The second phase would be a sample of the clients who would be interviewed to determine their frequency of use of services, whether they were homeless, whether they were also shelter users, and other questions. A systematic sample of clients may be a reasonable approach, with the sampling interval determined in advance based on the anticipated number of clients. The interviews would be conducted by trained staff who would also be responsible for obtaining the count of clients served and identification of the clients to be interviewed.

4. Street Component

Target population: The target population for the street component is the part of the street population that is not covered by either the shelter or the services components of the survey. The street population of homeless includes people staying overnight on the street, in abandoned buildings, parks, bus shelters, *etc.*

Sampling frame: The sampling frame would be an area frame consisting of a list of identifiable places where the street homeless might be found: sections of streets, parks, buildings, bus shelters, *etc*. Construction of such a frame should make use of people who work with the homeless to ensure that it is complete and also to identify areas where the homeless are more likely to be found. Using the knowledge of local experts, it is hoped that places on the frame can be divided into two, or possibly three categories: first to separate those areas where homeless are likely to be found from those where they are unlikely to be found, and then perhaps to further identify areas where they are almost certainly to be found.

Survey design: At least a couple of different options are being considered for sampling method. The first can be called the classical approach, in that it resembles the method used in some prior surveys of the street population, particularly the DC-MADS study. Stratification and sampling of places would be done independently for each of the categories on the frame, with a higher sampling rate in the areas where the homeless are more likely to be found. To simplify operations a two-stage sampling procedure might be used to select places, particularly in the lower likelihood areas. After selection of places it is assumed that interviewers would visit the places within a specified time period and attempt to interview everyone that they find.

The second approach being considered is based on snowball sampling, a type of network sampling. It would also start with a sample of places, but homeless people who are found would be asked where other homeless might be found, or perhaps where they were the previous night or nights. Methodological and operational feasibility of this approach is still being considered.

Reference period/Collection period: Should collection be done in a single night, or spread out over a number of nights? One possibility is to divide the sample of areas into a number of individually representative samples that could be covered on different nights. It would be possible for the same person to be in the sample on two different nights, since people can move between areas, and it will be necessary to decide how to treat such cases and to identify them in the field. Nevertheless, since each night's sample is representative, an estimate of the number of street homeless could be obtained for each

5

night, and the average of such estimates would be unbiased for the average size of the population. Unbiased estimates of the number of distinct individuals would be more problematic.

Data collection: In order to help in locating the homeless on the street, it is best to go looking for them at times when very few non-homeless would be there. Other surveys of this population have typically conducted data collection in the early morning hours, after most others have left the street and before morning business picks up – say between the hours of 2:00 and 5:00 a.m., for example. For reasons of personal security the interviewers also typically work in pairs. A similar approach to collection is proposed for this survey. Interviewers would approach all people that they find on the street in the sample areas to conduct the survey. Some screening questions would be used first to identify the homeless, and once identified the main survey questions could be asked. For improved response rates it would be better if at least one of the interviewers in each team is known and trusted by the street population.

The questionnaire would have to include some questions to determine if the respondent was on either the shelter or services frame, in order to avoid double counting of the total number of homeless. If direct questions are used to identify the overlap with other frames then the interview could stop as soon as that was determined, unless some characteristics of this overlap population were specifically of interest.

5. Global Issues

Before the implementation of any of the sample designs presented above, there are several issues affecting all three components that need to be addressed either through studies or field tests. Some of the issues are as follows.

1. Data quality, particularly for information related to identifying the frame overlaps and multiple chances to be included in the survey, is crucial. In shelters and soup kitchens co-operation of sampled individuals may be reasonable, since these people may already be used to answering administrative type questions in order to use the services; however, in some cases the quality of the responses may be in doubt. Some surveys, for example the DC-MADS survey, use a short exam to evaluate respondents ability to accurately answer the survey questions. For the street component or other types of services co-operation may also be an issue. For the street component even finding individuals may be a problem, since many of them may not want to be found.

2. Co-operation of individuals, particularly for the street component of the survey, may be increased through the use of incentives. Identifying appropriate incentives that will work well may be a challenge. Possible incentives include vouchers for meals or lodging, help in contacting other services or government programs that could help them, food, coffee, or cash. A number of different incentives may need to be tried.

3. A successful survey will require co-operation of the sampled shelters and service providers. Convincing reasons why it is in their interest to participate will need to be formulated and communicated effectively.

4. Media relations will also need to be planned.

5. The individuals covered by the three different sub-populations, and thus the sampling frames, will depend on the reference period of the survey. For example, more people will be in the services sub-population over a period of a week than in a single day; on the other hand, more of these people will also be in the shelter population.

6. As described in the introduction, there are some implications of having multiple frames (*i.e.* shelters, services, and street). Specifically, it is possible for an individual in the services population to also be in the shelter population, or people in the street population to also be in the services population. Similarly, someone who is in the HIFIS part of the shelter population may also be in the non-HIFIS part. These possibilities must be accounted for in order to calculate survey weights that are unbiased for the number of distinct homeless people. As noted earlier, this will have implications for the information that needs to be collected in the survey. There are two basic approaches: (1) ask direct questions of sample individuals to determine if they are also on other frames, and (2) form unique identifiers and estimate the frame overlaps based on the sample overlaps. The first approach an important advantage. First, with this approach it would be known for each sample element which frames they belong to, allowing analysts to study, for example, the subpopulation of the street population that do not use either shelters or services. A disadvantage of the second approach is that the estimates of the frame overlaps would be quite unstable, since the sample overlaps would tend to be quite small (being samples of samples) unless the sampling rates were very high. It is possible that the first approach could not be used with the HIFIS population; however, people interviewed in non-HIFIS components of the survey could be asked direct questions about their shelter use to determine if they are also part of the HIFIS population.

7. There is the additional difficulty that individuals may use more than one shelter or service, or the same one more than once. Thus it will be necessary to know all of the services that were used by each respondent, and how often they were used, during the survey period, since respondents might be sampled at any of the services that they use. However, it may be practically impossible to obtain this information, since some of the service uses during the survey period may actually occur after the interview. One possible way to get around this difficulty is to take frequency of use during a period of time before the interview as a proxy for frequency of use during the survey period. This information could be collected by asking the respondent about their use of services within a period of time before the interview.

8. Another collection issue that needs to be considered is whether it will be permitted for the same person to be interviewed twice, at two different shelters or services or at the same shelter or service on two different occasions. It is probably best that this not be allowed.

9. Similarly, it must be decided whether individuals who are also in the shelter population would be interviewed in the services survey. If a separate estimate of the number of distinct service users is needed then such people would need to be interviewed, at least to the extent of knowing their frequency of service use. Alternatively, people appearing in either the shelter or service samples could be asked about their participation in both the shelter and service systems, and the interview could then be terminated immediately once it is determined that the person has already been interviewed.

10. The structure of the interviews must be finalized. For example, the number of interviews per sampled service provider/service period combination, the length of the interview and how many interviewers will be necessary on an interviewing team must all be addressed.