PROBLEMS IN SAMPLING NOMADIC POPULATIONS

William D. Kalsbeek, University of North Carolina
Anne R. Cross, Westinghouse Health Systems

1. Introduction

Economic and social development in developing countries is partially dependent upon a better understanding of the nomadic segments in these societies. Studying nomads, however, has been impeded by the difficulty in obtaining scientifically valid samples of nomadic households when the research vehicle is the sample survey. The purpose of this paper is to briefly present several approaches that have been used to sample nomads and to describe in some detail the approach that was used for a survey in Somalia.

In 1980/81 the POPLAB project at the University of North Carolina assisted the Somali government in carrying out a demographic survey sample in the capital city, Mogadishu, and two agricultural regions nearby, in the southern part of the country. The survey gathered data about fertility, mortality and migration as well as background characteristics such as age, sex, literacy and marital status (Central Statistical Department 1981). Although the bulk of the survey effort was concentrated on interviewing the settled population, a smaller-scale survey covering nomads was also mounted.

According to standard definition (United Nations 1977 and American University 1977), "pure" nomads are those who depend entirely on their animals for their livelihood and who consequently move about in search of water and pasture and thus have no permanent residence. In Somalia, these pure nomads own an average of between twenty and several hundred camels as well as perhaps some cattle, sheep and goats. Semi-nomads, who engage in agriculture during part of the year, were treated as pure nomads in this study. Contrary to popular belief, nomads follow a fairly regular migratory pattern throughout the year. During the two annual dry seasons in Somalia, the herds are usually brought back to the home wells where there is a more or less reliable source of water. When the rains come, the nomads take their animals in search of green pasture where they often set up camps in the same place year after year.

The basic social unit of nomadic society is the household which usually consists of a single nuclear family. At certain times of the year, the family may be divided up, the men and older boys tending the camels while the women and children tend the cattle, sheep and goats. This custom of splitting up the household has major repercussions on sampling design as we will discuss later.

It has been estimated that over half of Somalia's population is nomadic. This unusually large proportion is undoubtedly due to the fact that the vast majority of the country receives less than 20 inches of rain annually and is unsuitable for cultivation. It is clear that nomads form an important sector of Somali society, and that development of adequate means by which to study them is crucial.

2. Review of Sampling Approaches

Five distinct approaches for sampling nomadic populations are briefly discussed by deGraft-Johnson (1979) based upon a more detailed review of methods by the United Nations (1977). These approaches specifically address the problem of conducting nomadic sample surveys in Africa, although there is little doubt that these methods could be applied in other areas of the world as well. Each of these approaches will be discussed below.

(1) In the group assembly approach officials in selected local administrative districts are asked to "round-up" all nomads in their districts at the time of the survey and convince them to assemble at designated assembly points. To the extent that nomads cannot be found and convinced to assemble, this approach will lead to nonresponse problems in the survey. Moreover double coverage is possible when the nomads are relatively mobile and when the survey data are collected over a long period of time so that nomads have a chance to appear at multiple assembly points.

(2) In the camp approach, a cluster sample of nomads is selected from a list of camps, each representing a cluster of nomads. This approach is feasible to the extent that a complete list of existing camps, with adequate description of their location, can be prepared. Unfortunately, this approach often has substantial coverage problems because a list of camps will often be incomplete to begin with or out of date by the time that data collection begins.

(3) The social structure approach utilizes the tribal nature of the nomadic population. The presumption here is that each nomad can be identified through a hierarchical arrangement of society wherein the tribe is the first layer. For sampling purposes the tribe then can be treated as a stratum (when the number of tribes is small) or as a primary sampling unit in a multi-stage sampling design (when the number of tribes is large). Social subdivisions of tribes are then selected in subsequent stages until a sample of the smallest social unit is selected, whereupon the final stage sample is selected from a list of nomads in the smallest social unit. The social structure approach has only been found to be feasible when there is a strong bond to tribal origins among nomads, when tribal and subtribal leaders can be easily identifed and enlisted, and when accurate household lists can be obtained from subtribal leaders.

(4) In the enumeration area approach, each nomad is linked to a parcel of land with recognized geopolitical boundaries. The enumeration area thereby represents a cluster of those nomads occupying the area at the time of the survey. A sample of enumeration areas is first selected, after which nomads are then identified and interviewed within each selected enumeration area. As with some of the other approaches, coverage may be an important problem with the enumeration area approach since an accurate canvass of even a small enumeration area may be difficult, especially when the nomadic population moves rapidly over a wide area (i.e., beyond the boundaries of the enumeration area).

(5) Perhaps, the most generally practical and technically sound approach to solving the problem of sampling nomads is the waterpoint approach, in which nomads are associated with those places

where they seek water for their animals or themselves. Waterpoints, therefore, represent clusters of those nomads who would seek water from them during the survey period. In general, the waterpoint approach calls for selecting a sample of waterpoints and then selectively interviewing nomads who use the sample waterpoints during some specified period of time.

Since the waterpoint approach was considered the most appropriate method for sampling nomads in Somalia, it will be described in more detail in the remainder of this paper.

3. Design Considerations in Sampling Nomads at Waterpoints

3.1 General Considerations

Several assumptions are made in designing a sample of nomads via the waterpoint approach. First, it is assumed that all nomadic households have a herd of animals that they tend. Second, most nomadic herds consist of some combination of cattle, camels, or sheep and goats. Third, during the dry season, the average length of time between watering is about two days for cattle, four days for sheep and goats, and eight days for camels. Fourth, the length of time between visits to waterpoints for a nomadic herd is determined by the amount of time between watering for the type of animal requiring watering most frequently. The type of animal in a herd which determines the frequency of watering will be called the herd's predominating animal. For example, if the herd for a nomadic household consists of cattle and camels, it is expected that this household will appear at a waterpoint about every two days, the average time between watering for the cattle, this herd's predominating animal. Finally, it is assumed that a more or less complete list of waterpoints can be produced for use as a sampling frame and that reasonably good measures of the number of household which obtain water at the waterpoint per day are available.

Suppose that sampling of nomadic households is done in two stages. In the first stage, a sample of waterpoints is chosen. For each selected waterpoint we select a period of time during which a portion of the nomadic households watering at the waterpoint are interviewed. Screening will refer to the process of determining which households encountered at the waterpoint will be interviewed. The overall probability of selecting a nomadic household at a waterpoint under this selection method is determined by four things: the probability of selecting the waterpoint, the length of the period during which data are collected at the waterpoint, the interval between watering for the household, and the number of groups into which the household has split (if any) at the time of the survey.

Given the above, there are basically two ways to proceed in designing a sample of nomads via the waterpoint approach. One is to manipulate the waterpoint selection probability and the length of the interviewing period at each waterpoint so that overall selection probabilities for all nomadic households in the population are equal. We will call this the epsem design (for equal probability of selection method). In the second design a procedurally simpler selection method will yield unequal selection probabilities for nomadic households. This method will be called the nonepsem design. We now compare the epsem and nonepsem sampling designs in greater detail.

We begin by defining an enumeration unit to be a nomadic household (or part of a split nomadic household) which obtains water for their animals at a waterpoint. An episode in which an enumeration unit approaches a waterpoint will be called a watering encounter. The following additional terms are defined for use in the remaining discussion:

- E_i = Average number of watering encounters per day at the i-th waterpoint.
- D = Total number of days spent screening for interviews at each sample waterpoint.
- S_{ij} = Number of days at the i-th waterpoint spent screening for enumeration units in whose herd the j-th type of animal is the predominating animal.
- W_j = Average number of days between watering for the j-th type of predominating animal.
- N = Total number of waterpoints in the population.
- n = Number of waterpoints selected in the sample.
- $\bar{m}$ = Prespecified average number of nomadic households interviewed per waterpoint.
- λ = Anticipated number of completed interviews per screened watering encounter (takes into account failure to meet screening criteria and survey nonresponse).

3.2 Sample Selection in the Epsem Design

The selection procedure leading to an epsem sample of nomadic households involves three steps. First, using E_i as the measure of size, a sample of n waterpoints is selected with probabilities proportional to size (see Cochran, 1977). The selection probability for the i-th waterpoint is thereby nE_i/E, where $E = \sum_{i}^{N} E_i$ is the total number of watering encounters per day.

Secondly, allowance must be made for the fact that nomad households have "naturally" unequal probabilities of selection at a waterpoint due to the differing watering intervals of their herds. This can be accomplished by varying the lengths of the interviewing period according to the predominating animals being herded so that, for example, the interviewing period for nomads with camels would be roughly twice as long as for nomads with sheep and goats.

To yield an epsem design, the length of the interviewing period at the i-th waterpoint for herds with the j-th type of predominating animal must be calculated as:

$$S_{ij} = \frac{W_j \bar{m}}{D\lambda E_i} \quad . \tag{1}$$

Screening periods must be designated for each type of predominating animal at each waterpoint. This requires that S_{ij} be determined for each of the

three types of predominating animals and then three periods of length S_{ij} must be selected. To limit the amount of screening time at each waterpoint, the following strategy for selecting three periods of length S_{ij} might be used. Suppose we let j=1 to refer to camels as the predominating animal in a herd, let j=2 refer to the sheep and goats, and let j=3 refer to cattle. Then $W_1=8$, $W_2=4$, and $W_3=2$ so that $S_{i1} > S_{i2} > S_{i3}$ and $S_{i1} > S_{i2} + S_{i3}$. At the i-th waterpoint a period of length $S_{i1}=D$ would be selected for all screening activities. Within this period, screening enumeration units with predominantly sheep and goats or cattle might be randomized separately as well. The resulting screening schedule might call for screening both camels and sheep and goats for a period of length S_{i2}, screening only camels for the next period of length $S_{i1}-S_{i2}-S_{i3}$, and then screening both camels and cattle for the last period of length S_{i3}.

To keep short the amount of time spent at each waterpoint in order to reduce the chances of multiple appearances at waterpoints, we wish for the average number of sample interviews per waterpoint ($\bar{m}$) to be kept small and for the waterpoint measures of size (E_i) to be large. To accomplish the latter, some of the smaller waterpoints might be combined with neighboring waterpoints. For example, to keep S_{ij} for camels less than two days with D=2, we prefer that

$$\frac{\bar{m}}{\lambda E_i} < \frac{1}{2} \qquad (2)$$

for all waterpoints.

The epsem characteristic of this design can be verified by treating the interviewing period of length, S_{ij}, as a random period in time. The probability of selecting an encounter unit with the j-th type of predominating animal, given that the i-th waterpoint has been selected, is thereby approximately S_{ij}/W_j. Considering the selection probabilities in both stages of sampling, the overall selection probability for each nomadic household with the j-th type of predominating animal would be

$$\left[\frac{nE_i}{E}\right] x \left[\frac{S_{ij}}{W_j}\right] = \left[\frac{nE_i}{E}\right] x \left[\frac{\bar{m}}{D\lambda E_i}\right] = \frac{n\bar{m}}{D\lambda E}, \qquad (3)$$

which is constant for all types of herds thereby implying an epsem design of nomadic households.

As a final step in ensuring an epsem design, the problem of multiple appearances by a nomadic household is avoided by using a so-called unique counting rule for screening, wherein a nomadic household represented by a watering encounter is interviewed only if the eldest woman 15-49 in a split nomadic household is part of the enumeration unit present at the watering encounter. Assuming that D is sufficiently small and that travel among selected waterpoints is ordered so that the same enumeration unit cannot appear in the sample more than one time, each nomadic household has but one chance of being interviewed in the survey when this unique counting rule is used.

The main advantage of the epsem design is that the equal selection probabilities make survey estimates easier to compute and more precise (see Hansen, et. al., 1953) resulting in less complicated analysis of survey results. One major disadvantage is that the survey interviewing team must possess suitable skills and training to be able to screen as well as to interview. Screening, as we have seen, requires that the screener (1) be aware of which kinds of herds are being screened at any given time and (2) accurately apply the unique counting rule. Correct application of the screening criteria may be too much to expect of the interviewing team in some surveys. Another disadvantage is that since a substantial proportion of potential respondents are screened out of an interview, interviewer productivity (i.e., measured as the number of interviews completed per hour) is suboptimal.

3.3 Sample Selection in the Nonepsem Design

Sample selection can be simplified somewhat and screening for interviews at selected waterpoints can be eliminated altogether by using a nonepsem sampling design. Sampling would be once again done in two stages but selection probabilities for nomadic households in the population would not be equal, thus requiring that special weighting factors be computed and used in analysis.

Selection of the sample of nomadic households in the nonepsem design proceeds as follows. An equal-probability sample of n waterpoints is selected from N total waterpoints in the first stage of selection. The first-stage selection probability for each waterpoint is therefore n/N. Then, a fixed interviewing time period of length D is identified for each selected waterpoint. An interview is requested of all nomadic households (regardless of the kind of herd) that water animals at selected waterpoints during the designated interviewing time period. During the survey interview, the following sampling information is obtained from each nomadic household: (1) the types of animals in the herd that was being watered at the time of the interview, (2) the length (in days) of the last interval between watering for each kind of animal in the households' herd, and (3) how (if at all) the household is currently split.

The overall selection probability for each nomadic household is roughly proportional to the ratio $\Sigma(W^*_{ik\ell})^{T_{ik}}$ where T_{ik} refers to the number of enumeration units into which the nomadic household (represented by the k-th selected enumeration unit in the i-th waterpoint) has been split at the time of the survey, and $W^*_{ik\ell}$ is the length of the last between-watering time interval for the predominating animal in the ℓ-th part of the ik-th enumeration unit. T_{ik} is called the "multiplicity" of the same household since it is the number of chances the household has to be selected in the sample. For example, a nomadic household split into two enumeration unit has two chances of being selected while a household which remains intact has only one chance of being chosen depend-

ing of course on the types of animals herded by each branch of the household. Thus, it is reasonable to expect that a household's selection probability is directly proportional to T_{ik}. On the other hand, the household's selection probability is inversely proportion to $W^{*}_{ik\ell}$ since the household with a longer period since it appeared at a waterpoint is less likely to be selected than a household with a shorter period since the last watering.

Three general implications of this nonepsem design should be noted. First, computation of weighting factors, while conceptually simple, is operationally burdensome especially when large samples are selected. Second, holding sample sizes fixed, the loss in precision of survey estimates is directly related to the variation of the weighting factors (see Kish, 1965), which is largely determined by the distribtuion of watering intervals for predominating herd animals in nomadic households. Finally, unless they are used to stratify waterpoints, measures of size are unimportant in this nonepsem design, which is an advantage when measures of waterpoint usage are of questionable quality.

4. Implementation of the Nonepsem Design in Somalia

A stratified nonepsem design was adopted in the POPLAB survey of nomads primarily because it was considered to be more operationally efficient than the epsem design. The first step in designing the sample for the survey of nomads in Somalia was the construction of a frame of waterpoints in the study area. To do this, teams were sent to local officials and nomad leaders to compile a list of the names, locations and measures of size of all major waterpoints servicing nomads. Wells located in villages and thought to service mostly villagers were excluded from the list. Although this waterpoint list was checked for completeness against previous lists, its accuracy was largely unknown. It should be mentioned that compiling the sampling frame of waterpoints for the study area took over two months and that replication on a national scale would be a major effort.

The list of waterpoints was next stratified by region and the number of reported watering encounters per day. A disproportionate stratified simple random sample of 60 waterpoints was selected with three times higher sampling rates in the Bay region where, the number of encounters per waterpoints was higher. All enumeration units approaching a selected waterpoint in the Bay region during an 8 hour period were interviewed. The interview period in all other selected waterpoints was 24 hours. Fieldwork was conducted by twelve interviewers, three supervisors and the senior field coordinator, all of whom were regular government employees and most of whom were men in their 20's and 30's. Because all the field staff had prior survey experience, training was confined to four days and emphasized practice interviews. Field staff were organized into three teams, each with a Land Rover and driver; and since many waterpoints were remote from villages and interviewing sometimes continued into the night, each team was equipped with tents, bedding, cooking supplies and food. In order to facilitate supervision and to maximize communication in an area somewhat smaller than the state of Maine, where roads are dirt if they exist at all and where telephone and telegraphs are virtually nonexistent, the three teams generally moved together from district to district.

Unfortunately, the survey in Somalia suffered seriously due to the weather . The survey commenced at the end of a severe drought and initially there was concern that a large proportion of waterpoints would be dry. This apprehension was soon totally overshadowed by the opposite concern, when after about ten days of fieldwork, the long rains came, over a month earlier than usual. Rain hampers interviewing in three ways: (1) nomads are no longer dependent on the waterholes since rainwater collects in gullies and puddles, (2) some waterpoints become inaccessible due to muddy roads; and (3) even areas not yet hit by the rains are evacuated by nomads migrating towards the rain and greener pastures. Thus, of the 60 waterpoints selected in the sample, half produced no interviews whatsoever. While drought or rain can hinder fieldwork among settled populations, either can virtually destroy a survey of nomads by wiping out one's access to respondents.

A more minor problem was the difficulty of identifying prospective respondents. Waterpoints can be busy places, bustling with women from villages collecting the family's daily water supply, people washing clothes and even bathing themselves. Although nomads usually have distinct hairdos and clothing, it was not always easy to know which group of people had brought which animals to water or to ensure that all nomads were interviewed. In addition, some wells had more than one opening sometimes up to one mile apart that required splitting up the interviewers to ensure complete coverage.

Even with the nonepsem sampling design, the waterhole approach to sampling nomads still results in much less efficient use of interviewers' time than in most designs for covering settled populations. This is mainly due to the fact that it is impossible to predict the number of nomads that will appear at a given waterpoint on a given day to water their animals. Rather than risk having too few interviewers, it was decided to err on the side of too many and since interviewers worked in teams, it was necessary to assign interviewers to waterpoints in multiples of four, resulting in a good deal of non-productive time for some interviewers. Thus, even before the onset of the rains, interviewers were averaging only about three interviews per day.

We are led from the review of existing procedures and our experience in Somali to conclude that sampling nomadic populations is possible though difficult to do well. Moreoever, we conclude that perhaps the most generally useful method for sampling nomads is the waterpoint approach since alternative methods seem to have a greater potential for nonsampling errors and operational difficulties. This is not to imply that the waterpoint approach is free of some important problems; however, we have noted some of the more important problems of the waterpoint approach earlier in the paper. First, selecting a statistically useful probability sample of nomads by the waterpoint approach requires a complete list of wa-

terpoints which is often difficult to produce. Second, when selecting an epsem sample, the required information on waterpoint usage may not exist and sophisticated interviewing staff needed for screening may not be available. Third, the problem of giving multiple chances of selection to split households or household which move rapidly among waterpoints can be solved but not without adding substantial complexity to sample selection or analysis. Fourth, herd composition and differential watering intervals also affects the selection probabilities for nomadic households. This problem, like the multiplicity problem, can be remedied by making suitable adjustments in sample selection or analysis. For both problems the remedy is only partial since the effects of multiplicity and differential watering intervals cannot be measured precisely. Finally nomadic surveys are often conducted in arid or semi-arid regions where travel is difficult. Selection of households to be interviewed and the interviewing process are therefore done in virtual isolation, thus making quality control measures difficult to implement.

REFERENCES

American University (1977), Area Handbook for Somalia, Foreign Area Studies, Second Edition, Washington, D.C.

Central Statistical Department, Ministry of National Planning (1981), Report on the Demographic Survey of Banadir, Bay and Lower Shebelle Regions of Somalia 1980-81, Mogadishu. Summarized in Summary Series No. 4 International Program of Laboratories for Population Statistics, Chapel Hill, N.C., December 1981.

Cochran, W. (1977), Sampling Techniques, Third Edition, Wiley and Sons, New York.

DeGraft-Johnson, K.T. (1979), "Question/Answer," Survey Statistician, NR. 1, International Association of Survey Statisticians, 22-23.

Hansen, M., Hurwitz, W., and Madow, W. (1953), Sample Survey Methods and Theory, Vol. I and II, Wiley and Sons, New York.

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

United Nations (1977), "Study on Special Techniques for Enumerating Nomads in African Censuses and Surveys," Paper presented at the Tenth Session of the Conference of African Statisticians, Addis, Ababa, 17-22 October 1977.