THE USE OF NETWORK SAMPLES IN ESTIMATING INCIDENCE OF MISSING CHILDREN

Seymour Sudman, University of Illinois at Urbana/Champaign

This paper describes the design of a study using network methods to estimate the incidence of missing children in the United States. (This study is scheduled to be fielded during the fall of 1986.)

We take as given the need and importance of carefully estimating the incidence of missing children. The allocation of resources, as well as the development of policies for agencies dealing with missing children, depends on reliable estimates of their numbers and characteristics. Unfortunately, such estimates are not easy to obtain. In addition to definitional problems that would be found in any study, there are several special difficulties in studying missing children:

a. The actual numbers of households in which a child is missing for any reason is small and becomes even smaller if one wishes to distinguish between alternative reasons for being missing and between population subgroups. Even very large samples may uncover too few cases to be sufficiently reliable. It should be noted, however, that rareness of the population is not unique to missing children but is found for many other populations that are important for policy evaluation.

b. Answering questions about missing children may be threatening to some respondents, thus leading to substantial underreporting. This would be the case if the child either is a victim of parental kidnapping or is a runaway, the two major reasons for a child being missing.

These problems may suggest that survey procedures should not be used, but the alternative methods are probably even less valid. The National Crime Survey obtains far higher levels of reported crime than are obtained from administrative records, and it is likely that careful surveys using multiplicity sampling and capture-recapture procedures will obtain better estimates of missing children than it is possible to obtain from administrative records or other sources.

In the typical survey, such as the National Crime Survey, respondents are asked either about only themselves or about all household members. For rare populations, the number located is small, often one or less per 100 contacts. Theoretically, there is no reason to limit the interview only to household members. Respondents could be asked about other persons, relatives, co-workers, neighbors, fellow members in organizations, friends, and acquaintances. To make the data useful, however, the respondent must be able to give reliable information about these additional persons and must also be able to report the size of the network so that it is possible to compute the probability of any individual being selected in the sample. If this can be done, it is possible to make unbiased estimates of the incidence of the rare population that are more reliable than simple household estimates.

As a simple example, suppose one wishes to estimate the population of all persons who are legally blind. One could ask respondents not only about all persons in their household but also about any brothers and sisters, regardless of where they live in the U.S. A person who has no brothers or sisters has only one chance of falling into the sample. This occurs if that person's household is selected. A person with two brothers and a sister living in three different households has four chances of falling into the sample. That person will be mentioned if either his/her own household or the household of any of his/her three siblings is selected. To compute unbiased estimates, the data must be weighted. The person with no siblings gets a weight of 1, whereas the person with three siblings plus him/herself gets a weight of 1/4.

It is also possible using these procedures to obtain sufficient information to locate the members of the rare population so that they can be interviewed directly. Thus, network procedures are used not just to measure incidence but also as an efficient and unbiased method for location.

A final use of multiplicity procedures occurs when the question is sensitive and respondents may be unwilling to report about their households. These respondents may be more willing to report about others whom they know outside their own households. Examples of topics where this might be the case are child beating and alcoholism. This may also be the case for missing children. In the case of sensitive questions, there may be ethical problems with asking respondents to report about others who can be identified, but there are no problems if the data are used simply for estimation purposes and not to locate the rare population.

Some users of survey data are concerned about multiplicity procedures because sampling variances are increased owing to the weights used to account for the differential selection probabilities. In other uses of this method, however, it has been found that there are only moderate increases in sampling variances while there are very large cost reductions. Thus, taking both cost and variance into account, multiplicity samples are much more efficient for rare populations than are standard household samples.

Others have been concerned about multiplicity methods because respondents do not give completely accurate information about persons in other households. It must be remembered, however, that reports about persons in the respondents' households are not perfect either. The question is whether there is differential accuracy of reporting, and if there is, what the magnitude of the difference is. In several applications mentioned later in the literature review, reporting about others in a network has been shown to be only very slightly less accurate than reporting about household members.

Since households with missing children are rare (if one uses any policy-relevant defini-

tion of missing) and since the topic may be sensitive, the use of multiplicity sampling seems promising. There are, of course, several key questions:

a. Will respondents be able to report accurately about missing children in other households?

b. If yes, what types of networks can be used to obtain accurate information about missing children? By network types we mean relatives, neighbors, co-workers, etc. Past research has indicated that as the network size increases and the frequency of contact decreases, reporting about other network members becomes less accurate. On the other hand, the larger the network, the greater the amount of information that is obtained. The optimum tradeoff between quantity and quality of information must be determined by empirical research.

Literature Review

The aim of multiplicity, or network, sampling is to spread the identification of members of the special population more broadly over the total population, thereby reducing the number of screening contacts needed (Sirken, 1970, 1972).

Linkages to close relatives have been used in multiplicity samples for several surveys of rare illnesses (e.g., Czaja et al., 1984; Sirken, Graubard, and McDaniel, 1978; Sirken et al., 1980), for a survey of births and deaths (Nathan, 1976), and for a survey of Vietnam era veterans (Rothbart, Fine, and Sudman, 1982). Linkages to neighboring households have been used in a survey of ethnic minorities (Brown and Ritchie, 1981; Snow, Hutcheson, and Prather, 1981) and in a pilot survey of home vegetable gardeners using sewage sludge (Bergsten and Pierson, 1982). Rothbart et al. (1982) and Czaja et al. (1984) found that with networks of relatives it is almost always possible to locate the members of the special population. Even if the informant does not have the complete address or telephone number, he or she can give the names of other relatives who will know the location of the member of the special population.

Other networks such as more distant relatives, co-workers, or members of the same church or other social organization may also be considered for use in locating members of special populations. Sudman (1985) has shown that for visible characteristics, such as physical handicaps, reports from co-workers and members of the same church or social group are reasonably accurate. As the groups become larger, however, the accuracy of reports of network members decreases, so that one is usually better off with smaller networks.

A wide range of procedures have been used to elicit network information. We give only a few illustrative examples. Perhaps the most detailed was obtained by Gurevitch (1961), who obtained information from respondents based on diaries that were kept for 100 days. Respondents were required to keep the diary with them at all times and to report all contacts. Many researchers have provided respondents with lists and asked them to identify individuals who are, for example, acquaintances, persons whom they know and who know them (Gurevitch and Weingrod, 1978), or persons with whom they talk about scientific problems (Friedkin, 1978). More commonly, particularly in kinship studies, no lists are available and therefore the respondents are asked for names (Bott, 1971; Boissevain, 1974). Distinctions may be made, as by Bott, between kin who are intimate (frequent visiting and mutual aid); effective kin who exchange Christmas presents or cards and are invited to each other's weddings and children's christenings; noneffective relatives who have no contact but have some knowledge of each other; and unfamiliar relatives about whom informants know nothing or virtually nothing. Adams (1968) asked about other relatives simply by asking for a number, although specifying location.

Erickson, Nosanchik, and Lee (1981) report a study of 43 bridge club members randomly split into three groups who were asked to identify long, medium, and short lists of fellow members. As the length of the list increased, there appeared to be some drop in the percentage identified. (Here the universe size was fixed.)

Excluding the literature on community elites, the studies that have validated network size information are rare. From the multiplicity estimation direction, the study by Nathan (1976) validated birth and marriage information among very close kin; similar validation was observed in reports of cancer patients by close kin (Sirken et al., 1981). Rothbart et al. (1982) compared reports of respondents about sons, brothers, and nephews who had served in Vietnam and demonstrated that aunts and uncles were substantially less accurate in reporting than were siblings, who in turn were slightly less accurate than parents.

Killworth and Bernard (1976, 1979) and Bernard and Killworth (1977) monitored teletype networks of deaf persons, ham radio operators, and office employees and then asked respondents to rank frequency of communication with each other. (See also Bernard, Killworth, and Sailer, 1980.) The results indicated poor correlations between rankings and logs or observations of contacts. It should be recognized that these communication events may have been of low salience.

Study Design

The purpose of this proposed research is to measure the accuracy with which specified networks report missing children. It appears that simply starting with probability samples of the general population would yield some cases, but not enough for a careful comparison of alternatives. For this reason, we propose an additional sample of known households with a missing child.

The source for such a sample is the Illinois State Police's ISEARCH files. The ISEARCH staff are already working with the Northwestern University researchers who are studying missing children and have indicated that they would work with the University of Illinois as well.

There might be a concern that releasing the names of households with missing children might

in some way infringe on the privacy of these households. We propose that any households selected because there is a known missing child be "salted" into a general population sample so that only the persons selecting the sample would know whether a specific household was part of the total population sample or part of the list sample. We would do this in any event to prevent interviewers from being influenced in their behavior by prior knowledge of a household's status.

Files will be kept under security at all times, and the name and address identifier files will not be located in the same place as the files containing survey results or sample status. The procedures that we plan to use have been developed from experience with a wide range of surveys on sensitive topics such as illegal behavior and cancer.

Location and Method of Study

The study will be conducted in the Chicago metropolitan area. We propose to conduct the study by telephone using the CATI (computer-assisted telephone interviewing) system of the Survey Research Laboratory. This system contains 24 interviewing stations, 16 in SRL's Urbana office and 8 in the Chicago office. A highly experienced group of interviewers is available.

Instrument

Questions asking respondents to identify persons outside the household who are members of a rare population, i.e., who have a missing child, come most naturally after questions have been asked about the household. For this purpose, we propose to use portions of the questionnaire being used in the study of missing children conducted by Northwestern University.

Sample

Past experience leads us to recommend that not more than one-third of the total sample be from lists of households with missing children; the other two-thirds of the sample would be a general population frame obtained from random digit dialing. For the purposes of this study, we would ignore the small percentage of households without telephones. We would plan to start with a sample of 600 households, of which 200 would be from lists.

Network Sample

Each respondent would be asked to provide the following network information:

a. Names of living parents of respondent and spouse (if in the geographical area under study)
b. Names of sons and daughters of respondent and spouse (in geographical area)
c. Names of brothers and sisters of respondent and spouse (in geographical area)
d. Names of next-door neighbors on either side
e. Names of three closest friends at work, if respondent is employed.

They would be asked if any of these specified persons has had a missing child within a specific time period. If any were reported, an interview would be attempted with that person to determine if the report was confirmed. In addition, approximately 500 persons identified in the network sample would themselves be sampled to determine whether they reported a missing child in the base household. This sample would be taken from the networks of respondents known to have a missing child. Approximately equal samples would be taken from each of the five network types listed above. In some households, the missing child would have returned home. We propose to conduct approximately 50 interviews with such children to provide information that not only would be useful in itself but also would provide important information for planning the capture-recapture study. Table 1 summarizes the proposed sample sizes.

TABLE 1

SAMPLE SIZES FOR MULTIPLICITY STUDY

Type		Size
Direct household sample:		600
Random digit dialing sample	400	
ISEARCH or other list sample	200	
Network sample (in households networked from list sample):		500
Parents of respondents (i.e., grandparents of missing child)	100	
Sons and daughters of respondents (i.e., siblings of missing child)	100	
Siblings of respondents (i.e., aunts and uncles of missing child)	100	
Next-door neighbors	100	
Co-workers of respondent	100	
Returned missing children		50
Total		1,150

It may be seen that the total sample is about double that of the primary sample, or 1,150 households. The decision to sample at this rate means that about 2-3 network members would be sampled from a household with a missing child. Additional sampling from the network would be possible but might create serious problems of contamination and network burden as the network members consulted each other.

Definition of Missing

We are aware of the conceptual as well as operational problems of defining a missing child. Especially for the network members, it would be necessary to think carefully about the time period during which the child was missing before asking the questions. Even if parents could report for very short periods, we would expect most network members to hear about a missing child only after some time had elapsed.

From an operational perspective, it would be necessary to determine an optimum recall period. At one extreme, one could ask if the

child had ever been missing, but that would lead to substantial memory errors about details. On the other hand, incidence rates for short periods would be low. We would make this judgment on the basis of consultations with the Northwestern University study staff, based on the results of their survey.

Analysis

There are four major classes of variables that we propose to measure in the analyses: sample cooperation, incidence, sampling variances, and costs.

a. Sample cooperation--Past experience has indicated that respondents are usually willing to provide information about their networks, although, as with all survey items, a few respondents may not answer the specific question. For some of the network types, there might be an unwillingness to report or a lack of information by respondents about missing children of others in that network. We do not expect this to be the case, but the first stage in the analysis would be to examine cooperation rates by various network types to see if there were statistically significant and practically important differences in cooperation rates.

b. Incidence--The key analysis would involve comparisons of the number of known missing children reported in the initially selected households and by network types. Past experience leads us to believe that the best reporting would be from the initially selected household, although even here there is likely to be some underreporting because of the sensitivity of the question. We expect the network reports to be a little less accurate than those from the initially selected household, but the actual level reported would determine how useful multiplicity methods would be for future research. As with cooperation, we would compare across network types to see if there were statistically significant and practically important differences in the numbers of missing children reported.

c. Sampling variance estimation--Sampling variance estimates would be computed for the various estimates of incidence in item b above. As mentioned earlier, the network estimates must be weighted by the differential probabilities of selection, which would increase sampling variance. On the other hand, information would be obtained about many more households, which would substantially reduce variances. The actual variances for the different network types cannot be predicted theoretically but could be determined from this research. They should be smaller than variances obtained using direct methods.

d. Costs--Network sampling procedures are not cost free. Obtaining network information adds to the cost of the interview, as do any location activities required to find a household identified as having a missing child. In this study, we would keep careful cost records so that these separate activities could be identified. It would then be possible to compare alternative multiplicity estimates with the standard procedures on the basis of cost-effectiveness, where effectiveness can be defined on the basis of either variance per unit of cost or mean-square error per unit of cost.

REFERENCES

Adams, B.N. (1968). Kinship in an Urban Setting. Chicago: Markham.

Bergsten, J.W. and Pierson, S.A. (1982). "Telephone screening for rare characteristics using multiplicity counting rules." Proceedings, Section on Survey Research Methods, American Statistical Association, 145-50.

Bernard, H.R. and Killworth, P.D. (1977). "Informant accuracy in social network data II." Human Communications Research, 4, 3-18.

Bernard, H.R, Killworth, P.D., and Sailer, S. (1980). "Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data." Social Networks, 2, 191-218.

Boissevain, J. (1974). Friends of Friends: Networks, Manipulators, and Coalitions. New York: St. Martin's Press.

Bott, E. (1971). Family and Social Network: Roles, Norms, and External Relationships in Ordinary Urban Families. 2nd ed. New York: Free Press.

Brown, C. and Ritchie, J. (1981). Focussed Enumeration: The Development of a Method for Sampling Ethnic Minority Groups. London: Policy Studies Institute and Social and Community Planning Research.

Czaja, R., Warnecke, R.B., Eastman, E., Royston, R., Sirken, M., and Tuteur, D. (1984). "Locating patients with rare diseases using network sampling: Frequency and quality of reporting." In Health Survey Research Methods: Proceedings of the Fourth Conference on Health Survey Research Methods, 1982. Publication No. (PHS) 84-3346. Hyattsville, Md.: National Center for Health Services Research, pp. 311-24.

Erickson, B.H., Nosanchik, T.A., and Lee, E. (1981). "Network sampling in practice: Some second steps." Social Networks, 3, 127-36.

Friedkin, N.E. (1978). "University social structure and social networks among scientists." American Journal of Sociology, 83, 1444-65.

Gurevitch, M. (1961). The Social Structure of Acquaintanceship Networks. Cambridge, Mass.: MIT Press.

Gurevitch, M. and Weingrod, A. (1978). "Who knows whom? Acquaintanceship and contacts in the Israeli national elite." Human Relations, 31, 195-214.

Killworth, P.D. and Bernard, H.R. (1976). "Informant accuracy in social network data." Human Organization, 35, 269-86.

Killworth, P.D. and Bernard, H.R. (1979). "Informant accuracy in social network data III: A comparison of triadic structure in behavioral and cognitive data." Social Networks, 2, 19-46.

Nathan, G. (1976). "An empirical study of response and sampling errors for multiplicity estimates with different counting rules." Journal of the American Statistical Association, 71, 808-15.

Rothbart, G.S., Fine, M., and Sudman, S. (1982). "On finding and interviewing the needles in a haystack: The use of multiplicity sampling." Public Opinion Quarterly, 46, 408-21.

Sirken, M.G. (1970). "Household surveys with multiplicity." Journal of the American Statistical Association, 65, 257-66.

Sirken, M.G. (1972). "Stratified sample surveys with multiplicity." Journal of the American Statistical Association, 67, 224-27.

Sirken, M.G., Graubard, B.I., and McDaniel, M.J. (1978). "National network surveys of diabetes." Proceedings, Section on Survey Research Methods, American Statistical Association, 631-35.

Sirken, M.G., Royston, P., Warnecke, R., Eastman, E., Czaja, R., and Monsees, D. (1980). "Pilot of the national cost of cancer care survey." Proceedings, Section on Survey Research Methods, American Statistical Association, 579-84.

Sirken, M.G., Royston, P., Bercini, D., Czaja, R., and Eastman, E. (1981). "Completeness of enumeration of persons with cancer in health surveys." Proceedings, Social Statistics Section, American Statistical Association, 223-27.

Snow, R.E., Hutcheson, J.D., and Prather, J.E. (1981). "Using reputational sampling to identify residential clusters of minorities dispersed in a large urban region: Hispanics in Atlanta, Georgia." Proceedings, Section on Survey Research Methods, American Statistical Association, 101-6.

Sudman, S. (1985). "Experiments in the measurement of the size of the social networks." Social Networks, 7, 127-51.