Articles

Sampling Rare and Elusive Populations

SEYMOUR SUDMAN, MONROE G. SIRKEN, CHARLES D. COWAN

The sampling of rare and elusive populations is difficult because the costs of locating such populations are substantial and can exceed actual interviewing costs. There are efficient probability methods that have been developed recently that reduce these costs. If the special populations are geographically clustered, efficient sampling involves the rapid location of segments in which no members of the special population are located with the use of Census data, telephone screening, or incomplete lists. Populations that are not geographically clustered can be located by network sampling and use of large previously gathered samples. Characteristics of mobile populations such as the homeless can be estimated by capture-recapture methods.

THE SAMPLING OF GENERAL POPULATIONS OF INDIVIDUALS or households has become well understood (*1–3*). In recent years, however, much of social research has dealt with special populations; examples include (i) racial and ethnic groups including blacks, Hispanics, Vietnamese, or Russian immigrants to the United States and guest workers in Europe; (ii) persons with incomes above or below a given amount such as the very rich or poor; (iii) persons with an illness such as cancer or cystic fibrosis; (iv) households with a missing child; (v) disabled scientists; (vi) black Vietnam war veterans; (vii) homeless persons; or (viii) recreational fisherman and their catches.

It may be seen that many special populations are rare and difficult to locate unless lists are available. If lists are available sampling is easy, but, for most rare populations, lists do not exist or are unavailable to researchers. In such cases, screening of the general population will be necessary, and the costs of screening can equal or far exceed the actual costs of interviewing. Some researchers with limited resources may be inclined to throw up their hands when undertaking studies of rare populations and to use ad hoc convenience samples. Although such convenience samples may sometimes be adequate for exploratory research, they are totally inadequate for making careful estimates about the special population.

Efficient probability methods of sampling rare and elusive populations produce useful population estimates at substantial reductions in cost (*4, 5*). With probability methods, every unit in the population has a known nonzero probability of selection, and it is possible to measure sampling variances. In this article, we discuss first efficient methods when the rare populations are geographically clustered (the first two examples given above); second, we outline procedures that may be used more generally, but especially for populations that are not geographically clustered (examples iii through vi above); and finally we discuss the sampling of rare mobile populations (examples vii and viii).

S. Sudman is Walter A. Stellner Professor of Marketing and Research, Survey Research Laboratory, University of Illinois at Urbana-Champaign, Champaign, IL 61820. M. G. Sirken is associate director for research and methodology, National Center for Health Statistics, Hyattsville, MD 20782. C. D. Cowan is chief statistician, Center for Education Statistics, U.S. Department of Education, Washington, DC 20208.

Geographically Clustered Samples

Standard cluster sampling methods reduce survey costs because multiple interviews are conducted in compact geographic segments. At the same time, cluster samples increase sampling variance in comparison with simple random samples of the same size. The increase in sampling variance is an increasing function of the sample size of the cluster and of the homogeneity of elements within the cluster.

Many geographically clustered special populations are not spread evenly across the United States, but are found in a limited number of states, cities within those states, and neighborhoods and blocks within those cities. There is a large fraction of total geographic segments in which no members of the special population are located. The standard cluster procedures in such a case require large numbers of screening visits or calls in these zero segments.

If the zero segments are known in advance from Census data or other sources, substantial cost savings are possible by elimination of the screening visits to the zero segments. Frequently, however, zero segments are not known in advance; Sudman (*6, 7*) has discussed optimum procedures when this is the case.

The efficient elimination of zero segments through use of one (or a few) screening contacts per segment can substantially reduce screening costs, particularly if the proportion of zero segments is high. The widely used method for improving the efficiency of random digit dialing telephone procedures described by Waksberg (*8*) may be adapted for special populations with cost savings.

The procedure requires that initially a single unit be screened within a geographic segment. If that unit is a member of the special population, additional screenings are conducted in the segment until a predetermined cluster size $k + 1$ is reached.

Among nonzero clusters, the probabilities of selection are N_i/S, where N_i is the number of special population units in segment i, and S is the size of the special population. To obtain a self-weighting sample, the sample size is set at $k + 1$. The overall probability of selection of a member of the special population in a cluster is

$$\left(\frac{N_i}{S}\right)\left(\frac{k+1}{N_i}\right) = \frac{k+1}{S}$$

It is possible to compute an optimum value of $k + 1$ that minimizes the sampling variance for a given cost. In many applications this value of $k + 1$ is between 5 and 10.

Substantial cost savings in surveys of about 70% or more are possible with this procedure when t, the proportion of the total population in segments with no special population members, is around 0.9 and π, the proportion of the special population to the total population, is correspondingly low. On the other hand, there is no advantage when t is less than 0.5 or 0.6 and π is greater than 0.2.

The efficiency of this method depends to a lesser extent on the relative costs of interviewing and screening and on the homogeneity (the intraclass correlation coefficient) within clusters as would be the case in standard cluster sampling. Figure 1 shows the relative data collection costs of optimum screening in comparison with simple random sampling. Here it is assumed that the screening costs are half the interviewing costs and that ρ, the measure of homogeneity, is 0.01, a typical value. Similar patterns are observed for other costs and values of ρ. An example here illustrates the effectiveness of geographic screening. It is assumed that telephone sampling and interviewing are used and that virtually all eligible households have telephones.

Telephone Screening of Black Households by Random Digit Dialing

Suppose one wishes to select a national sample of 1200 black households, as did Czaja and Blair (*9*). They estimated that black households are approximately 3% of all working U.S. telephone numbers and that about 70% of all working banks of 100 telephone numbers have no black households. Groves and Kahn (*10*) estimated that 65% of all telephone banks consist solely of nonworking numbers. Thus $t = 1 - (0.35)(0.30) = 0.9$ and $\pi = 0.03$.

The interviewing cost is estimated to be \$10, and the screening cost is estimated as \$2. For this example, assume that ρ, the estimate of homogeneity, is 0.05. Then, the optimum value of $k + 1$ is 8, and the field cost of the optimum design is \$29,000 as compared to \$68,000 for an unclustered sample with equivalent variance.

The Use of Incomplete Lists

Even incomplete lists may be useful in identifying areas where the special population is located. In the simplest case, assume that a random (or systematic) sample of starting points is chosen from the list and that field screening continues from each starting point until k additional eligible units are located in the cluster. It is evident that this procedure is almost identical to those just discussed.

A cluster will have an initial probability of selection proportional to the number of units of the special population in it that are on the list. Although the definition of the cluster is somewhat arbitrary, it must be made in advance before the sample is selected. The telephone exchange of the unit selected from the mailing list would be a natural cluster for samples for which mailing lists and telephone interviewing are used. For face-to-face interviewing, either the block or the zip code would be a natural cluster. Then, the sampling rate within the cluster is inversely proportional to this probability, so that the ultimate sample is self-weighting.

The sample is biased, however, if there are geographic clusters with eligible units, but none of them appear on the list. These clusters have no chance of selection. It is possible to measure the undercoverage from such a procedure if there is an independent estimate of the total size of the special population: One, estimate from the list the number of nonzero clusters and from the screening the average number eligible per nonzero cluster. Two, the product of the two is an estimate of the number of persons in the special population who have a nonzero probability of selection with starting points from the list. Three, the difference between the known size of the special population and two is the estimated undercoverage.

Dual Frame Methods

An alternative procedure for use of an incomplete list is to combine a sample from such a list with a sample of the total population. The total population sample is unbiased, but inefficient. The list sample is biased, but efficient. Hartley demonstrated that efficient, unbiased estimates are possible by combining the estimates, a procedure he called dual frame estimation (*11, 12*).

Let T be the size of the specified population not on the list, $\overline{X}_T$ the estimate of a mean value of a characteristic based on the sample from the total population not on the list, S the size of the list and $\overline{X}_S$ the estimate of the mean based on those in either sample who are on the list. Then, an estimate of a total X is given by $X = T\overline{X}_T + S\overline{X}_S$. It is possible, with the use of the kinds of cost functions that have been discussed, to minimize the variance of dual frame estimates by optimum allocations between the two frames.

Variations in Density of Special Population in Nonzero Clusters

The situation in which the special population is unevenly distributed among the nonzero clusters would be likely to occur with ethnic groups where most members live in a few geographic area clusters with high proportions of the population but others are thinly spread over the remaining nonzero clusters. Such information might be available from earlier screening or Census data or by asking the first contacted households to estimate it.

Telephone screening. Assume first that the nonzero clusters have been identified and categorized into strata when π_j is the proportion of the special population to the total population in stratum j. With telephone screening, no clustering is required, and the cost per case in the jth stratum is

$$C_j = CI + CS/\pi_j$$

where CI is the interviewing cost and CS the screening cost. The optimum allocation procedure would be to sample from the strata at rates inversely proportional to the square roots of costs. Thus, the relative rates in strata A and B would be

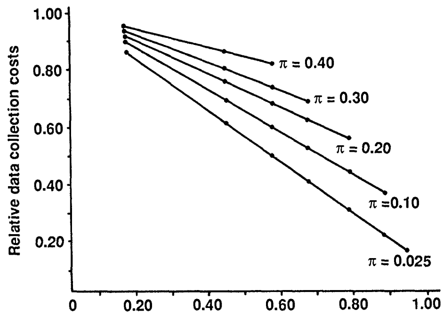
$$\frac{r_A}{r_B} = \left(\frac{CI + CS/\pi_B}{CI + CS/\pi_A}\right)^{1/2}$$

Kalton and Anderson (*4*) give tables that show that substantial gains from disproportionate stratification accrue only when π_B is much greater than π_A, and when a high proportion of the special population—that is, 80% or more—is in stratum B.

General Methods

When the rare population is not geographically clustered, other methods must be found for reducing screening costs. Among the methods that have been successful are the use of large previously collected samples or the addition of screening questions to ongoing or future surveys of the general population being conducted for other purposes. For example, Greeley and Rossi (*13*) used a sample from a large previous study of adult education activities to locate Catholics for their study of Catholic Americans; Harris *et al.* (*14*) used a prospective screening of victims of serious accidents by adding a question to a large number of omnibus type surveys until sufficient victims were located. The National Survey of Family Growth was based on a sample of women in the reproductive age groups previously interviewed in the National Health Interview Survey (*15*).

Fig. 1. Relative data collection costs for optimum screening versus unclustered. t is the proportion of total population in segments with no special population members; π is the proportion of special population to total population.



Network Samples

In the past few years increasing attention has been given to the use of network samples for locating and measuring the size of rare populations. The basic idea is simple—to increase the amount of information obtained during a screening interview.

In the typical survey, households are selected by a procedure that results in each residence having an equal probability of selection, that is, a self-weighted sample. However, usually one is more interested in individual than household estimates. Consequently, subsequent to household selection a household listing is obtained from one household member and an individual is selected from each household by random selection. When a special population is studied, only household members meeting the criteria would be included. For the sample to be unbiased, it is necessary that the selected interviewee be weighted by the number of other eligible persons living in the household. Otherwise, of course, the sample would be biased toward those living in smaller households.

Network sampling (or multiplicity sampling as it sometimes is called) expands on this procedure. Respondents can be asked about relatives, neighbors, co-workers, and members of their social groups living outside their own household.

Of course, weighting is necessary for someone selected by network sampling. For the sample to be unbiased, people who are eligible to be reported by all n members of a network must be weighted by the reciprocal $1/n$. This weighting provides what is called a multiplicity estimate. An example may be helpful. Suppose we define a sibling network as consisting of all households where the respondent or a sibling lives. Someone without brothers and sisters could be selected only if that person's household was selected. Someone with three brothers and three sisters, all in separate households would have 3 plus 3 plus 1 chances of selection or 7 in total. That person would get a multiplicity weight of 1/7.

A Brief History of Network Sampling

Much of the early work on network sampling was undertaken by Sirken and his associates at the National Center for Health Statistics (*16–23*). Fishburne and Cisin (*24*) employed network sampling in a survey on drug abuse; Nathan (*25*) in a survey of marriages and births; and Schmelz *et al.* (*26*) in a study of births and deaths in Israel. Sirken and Goldstein (*27*) proposed the method in surveys of Jewish populations, and Sirken *et al.* (*28*) used it in a study of diabetes.

There are clear cases in which network sampling has proved to be an efficient alternative. Rothbart *et al.* (*29*) describe the use of the approach to locate Vietnam War veterans, oversampling for black veterans. Parents and siblings were more accurate in their reporting of veteran status than aunts and uncles. Despite this limitation and the costs of locating referred interviewees, the total cost to obtain the required sample was about half that estimated for sampling by conventional screening procedures.

Czaja *et al.* (*30*) used multiplicity methods to locate cancer patients. They defined the network to include siblings and children outside the patient's home. Network sampling, according to their estimates, provided a much more efficient way of locating cancer cases than would have been obtained by conventional means. They concluded that unless total sample size is very large or the medical condition of interest has a prevalence considerably greater than that for cancer, network sampling is the preferred method.

Most network samples have used relatives as informants, but networks of close neighbors have been used by Brown and Ritchie (*31*) for a study of ethnic minorities and in a pilot study of home vegetable gardeners who used sewage sludge by Bergsten and Pierson (*32*). Sudman and Sirken (*33*) have recently proposed a study of work networks to measure the incidence and characteristics of disabled scientists in the workplace.

The Assumptions of Network Sampling

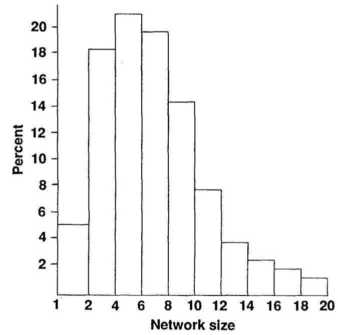
It is obvious that a prerequisite for any network sampling method is that network informants report accurately about all persons in their network. It would be unrealistic to expect perfect reporting from network informants since it is already known that self-reporting of one's own characteristics within the household is not perfect. What one would want would be network reporting that is not substantially worse than that obtained in standard household samples. Thus Czaja *et al.* (*30*) matched the reported cases of cancer obtained from the households and network informants to a cancer registry used for validation. There was about 90% agreement between patients' household reports and registry data and 83% agreement between sibling and parent network reports and the registry.

The quality of network reporting depends on the current social visibility of the characteristic being screened and the nearness of the relationship. Thus, a physical disability would be more likely to be known than would a behavior such as alcoholism or income tax evasion, although some network members would be aware of even these behaviors. Close relatives such as parents and children would be more likely to be aware of any rare characteristics than would be more distant relatives such as aunts, uncles, and cousins as seen in the network study of Vietnam veterans (*29*).

A recent example of a rare population not well reported by network informants was households with a missing child (*34*). Most missing children are runaways and many are gone for very short periods of time. In these less serious cases, the parent or guardian may well not discuss the event even with close relatives, neighbors, or co-workers. We found overall agreement between informants and the household with the missing child to be poor. For a small number of cases known to be kidnappings, however, network informants reports agreed with the household reports about 75% of the time. Whether this level of accuracy is sufficient would depend on the precision needed in a study. These negative results are reminders that network methods must be tested and validated before adoption.

Less obvious is the need to obtain an accurate estimate of the network size, either from the network informant or the person with the rare characteristic. This is very straightforward when the network consists only of close relatives, but becomes more difficult when more distant relatives are included and even more problematic when networks of neighbors or co-workers are used. In some recent work, however, Sudman (*35, 36*) has demonstrated that good estimates, on average, can be obtained from respondents about the

Fig. 2. Distribution of relative-network size from the study of national access to medical care (*38*).



number of people whom they know at work, although some random measurement error is observed. These estimates are better than those for church groups which is not surprising since work groups have more frequent and intimate contacts. Information about next door neighbors is also of generally high quality, but information about more distant neighbors on the same block is much less accurate.

In some cases, one merely wishes to estimate the size of the rare population, but in most cases, one is interested in locating the members of the rare population so that they can be interviewed. It will not always be the case that network informants will know the addresses or telephone numbers of the members of the rare population, or be willing to give them, but Rothbart *et al.* (*29*) found that it was usually possible to locate them by contacting other members of the network until finding someone who had complete location information.

Cost and Efficiency of Network Procedures

Because it is necessary to ask additional screening questions and to spend resources locating members of the rare population, network sampling costs slightly more per screening call than standard procedures. However, in cases where the method has been used, this cost is more than compensated by the reduction in sampling variance. If one is simply making an estimate of the rare population size then there would be no additional costs of locating the person identified by the network informant.

For estimating population sizes, the increased information obtained from networks is diminished to some extent by the increased variance caused by variability in multiplicity weights. Nevertheless, Nathan (*25*) showed that for fixed costs a network sample of mothers and sisters had a sampling variance about half as large as that from a conventional sample in a study that was designed to estimate number of births.

If one looks at the total cost of locating and interviewing members of rare populations, a network sample of size *n* has generally been found to be only about half as costly as a traditional sample of the same size. This comparison is somewhat misleading since the sampling variances for the network samples are larger than for standard samples of the same size because of variability in the weights.

Sudman and Freeman (*37*) have estimated the increased variances from network sampling in a large national telephone study of access to health care (*38*). Network sampling was used to increase the sample sizes of respondents with ten major chronic illnesses so that each of these groups might be studied separately. For this study, the network included parents, grandparents, children, and siblings of the adult respondent and spouse in the randomly selected household. Figure 2 shows the distribution in network size. The variability in this distribution leads to the variability of the weights and the increased sampling variances. In this study the ratio of the sampling error for the network sample as compared to a standard household sample of the same size was 1.25. Similar results were found by Rothbart *et al.* (*29*). It is clear that, for these studies at least, the net effect of network sampling is to substantially reduce the total sampling error since the effects of the larger network sample for the same costs are greater than the effects caused by variability in weights.

Multiplicity samples do not always lead to an increase in sampling variance because of variance in weights. Spaeth (*39*) has been studying hierarchical structure in organizations. In such studies one wishes to sample work organizations and supervisors with probabilities proportionate to the number of persons employed. In network sampling this is what happens. The probability of a supervisor being selected is proportional to the number of persons who report directly or indirectly to that organizational position.

Capture-Recapture Methods

Capture-recapture is a technique used to estimate the size of populations that are difficult to find and count, or populations that are in motion and cannot be counted all at one time. The technique was originally developed for use in counting populations of animals or fish, and has been extended to use with nomadic or mobile human populations (*40–43*). Capture-recapture is also commonly used as an evaluation technique for evaluating the completeness of coverage in a census or other form of enumeration that is supposed to represent the totality of a population (*44*).

The technique requires obtaining two or more independent observations on the same population. Most commonly for human populations at least one observation is based on a complete enumeration, though this is not necessary. The observations need to be taken at approximately the same times, or based on different sources that represent approximately the same population. The researcher needs to know only three things to make an estimate of the population size N_T : N_1, the number of persons observed at the first time (or in the first source); N_2, the number of persons observed at the second time (or in the second source); and M, the number of persons observed at both times (or in both sources). Table 1 shows how these counts are tabulated. An estimate of the number of members in the whole population, N_T, is $N_T = (N_1 \times N_2)/M$.

Observations are taken on the population to be studied in such a way that all members of the population have an equal chance of being observed in one capture (or equal chance of membership in one source). Those captured are tagged in some way so that at some later time the researcher can determine whether the individual has been observed earlier. With human populations this usually means collecting identifying information like name, address, date of birth, race, sex, and any unique identification, like social security number.

The basic concept of capture-recapture is very simple; the implementation of capture-recapture is difficult because the assumptions necessary to make the technique work often do not hold, or at least they do not hold without some intervention by the researcher. In deriving the estimator above, several assumptions are needed. The first has already been mentioned: each individual in the population has to have the same probability of capture during the observation period, though this probability does not have to be the same in period one and period two (or both sources). This assumption is frequently violated as different individuals can be missed in the counting process for very different reasons. The best example of this comes from the evaluation of the coverage of the 1980 Census, where the coverage rates varied for different race or ethnic groups. To deal with this problem, the researcher must stratify the counts from each source, making separate population estimates for each

Table 1. Observations for capture-recapture estimation.

First observation	Second observation		
	Captured	Not captured	Total
Captured	M		N_1
Not captured			
Total	N_2		N_T

group and then summing across these groups to get an estimate of the total.

A second major assumption necessary to the statistical model is that observing an individual at one time has no effect on the event of observing the individual at a second; this is the assumption of independence between counts. With more than two observations, this assumption can be relaxed to allow for correlation between observations as long as independence holds among all sources of information simultaneously. This assumption rarely holds for the case where only two sources are used for a human population, since being missed in one source, like a Census, is usually correlated with being missed in the second source, like a follow-up sample survey.

Even when one source is a census, and the second source is a carefully done sample, one often finds that certain types of persons are routinely missed in both sources. The homeless, migrant families or persons, families on vacation, and other similar groups that are hard to target because of their mobility or lack of a permanent address are good examples of groups that may be missed in both sources for similar reasons.

A third assumption, stated indirectly earlier, is that the population being studied does not change in composition or size during the elapsed time between observations. This assumption leads to the model being termed a "closed population model." If the population studied changes between captures, then the probability model used to make estimates is dealing with two population sizes, and so the model will not work. There is a procedure which can be used for a population that is changing in size, called an "open population model," but it requires at least four captures to be conducted since the model also has to estimate birth and death rates for the components of population change. Population changes in composition can also lead to biased estimates, unless population subdomains are estimated separately (*44*).

References have been made both to captures made as successive observations from the same source and also to captures that are contemporaneous, coming from different sources, such as administrative record lists which could be matched to a sample survey or other data collection activity. In this case, the assumption is made that each individual can be identified specifically and that each has the same chance of getting on the administrative record list.

Some populations are difficult to count not only because they are rare, but because it is difficult to determine whether someone does or does not belong to the population. An excellent example of this sort of special population is found in studies that attempt to count or describe the homeless (*45*, *46*).

In block count studies of the homeless, where a team of enumerators canvasses a city block to determine the number of homeless persons residing there, enumerators often have great difficulty identifying who is homeless. This problem is exacerbated by two factors; first, the homeless fend for themselves quite well in some cases and are able to get support like clean clothes through shelters, missions, and welfare agencies. The interviewing team cannot always tell whether a person is homeless just by observing an individual. Second, some homeless persons are only temporarily displaced, having recently lost a job for example, and so are only temporarily without support, and so would transit in and out of the population between successive counts. Third, some homeless are also vagabonds and travel to different places in different seasons.

All of these factors, as well as refusal to respond to the survey introduce a great deal of indeterminancy to the modeling process and make estimation difficult.

Summary

Some may be surprised by the attention that has been paid to costs in this paper, but efficient use of scarce resources is at the heart of all sampling techniques and is critical for rare populations. Even among rare populations there are varying levels of difficulty.

Rare populations that are geographically clustered are the easiest to sample of this difficult group. Once the zero segments have been eliminated, or samples with low levels of the eligible population sampled at a lower rate, costs are substantially reduced. Where there is little or no geographic clustering, obtaining data from social networks outside the household substantially increases the number of members of the rare population reported for only minor increases in costs. The key assumption is the network informants are able to provide the required screening information.

Most difficult of all are the rare mobile human populations. As was pointed out, fairly strong assumptions are necessary before estimates of the population size can be made with capture-recapture methods. The problems are far from solved at the present time, but, given the current interest in capture-recapture methods, it is reasonable to expect that new theoretical developments and applications of procedures will be forthcoming in the next decade.

REFERENCES

1. M. H. Hansen, W. N. Hurwitz, W. G. Madow, *Sample Survey Methods and Theory* (Wiley, New York, 1953).
2. W. Cochran, *Sampling Techniques* (Wiley, New York, 1963).
3. L. Kish, *Survey Sampling* (Wiley, New York, 1965).
4. G. Kalton and D. Anderson, *J. R. Stat. Soc.* **149**, 65 (1986).
5. S. Sudman and G. Kalton, *Annu. Rev. Sociol.* **12**, 401 (1986).
6. S. Sudman, *J. Am. Stat. Assoc.* **67**, 335 (1972).
7. ——— *J. Mark. Res.* **22**, 20 (1985).
8. J. Waksberg, *J. Am. Stat. Assoc.* **73**, 40 (1978).
9. R. Czaja and J. Blair, *Public Opin. Q.* **46**, 585 (1982).
10. R. M. Groves and R. L. Kahn, *Surveys by Telephone: A National Comparison with Personal Interviews* (Academic Press, New York, 1979).
11. H. O. Hartley, *Proc. Soc. Stat. Sect. Am. Stat. Assoc.* (1962), pp. 203–206.
12. ——— *Sankhya C* **36**, 99 (1974).
13. A. Greeley and P. H. Rossi, *The Education of Catholic Americans* (Aldine, Chicago, 1966).
14. D. R. Harris *et al.*, *Compensation and Support for Illness and Injury* (Macmillan, London, 1982).
15. National Center for Health Statistics, *Linking the National Survey of Family Growth with the National Health Interview Survey* (Vital and Health Statistics, Government Printing Office, Washington, DC, 1987) series 2, no. 103.
16. Z. W. Birnbaum and M. G. Sirken, *Design of Sample Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates* (Vital and Health Statistics, Government Printing Office, Washington, DC, 1965) series 2, no. 11.
17. M. G. Sirken, *J. Am. Stat. Assoc.* **65**, 257 (1970).
18. ——— and P. N. Royston, *Proc. Soc. Stat. Sect. Am. Stat. Assoc.* (1970), p. 361.
19. M. G. Sirken, *J. Am. Stat. Assoc.* **67**, 224 (1972).
20. ———, *Biometrics* **28**, 869 (1972).
21. ———, in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (Univ. of California Press, Berkeley, 1972), p. 135.
22. ——— and P. S. Levy, *J. Am. Stat. Assoc.* **69**, 68 (1974).
23. P. S. Levy, *ibid.* **72**, 758 (1977).
24. P. M. Fishburne and I. Cisin, *National Survey on Drug Abuse: Main Findings* (National Institute of Drug Abuse, Washington, DC, 1980), no. 108.
25. G. Nathan, *J. Am. Stat. Assoc.* **71**, 808 (1976).
26. V. O. Schmelz, G. Nathan, J. Kenvin, *Multiplicity Study of Births and Deaths in Judea-Samaria and Gaza Strip-North Sinai* (Technical Publications Series, Central Bureau of Statistics, Jerusalem, 1977), no. 44.
27. M. G. Sirken and S. Goldstein, *Proceedings of the Sixth World Congress of Jewish Studies* (Hebrew University, Jerusalem, 1973).
28. M .G. Sirken *et al.*, *Proc. Soc. Stat. Sect. Am. Stat. Assoc.* (1975), p. 659.
29. G. S. Rothbart, M. Fine, S. Sudman, *Public Opinion Q.* **46**, 408 (1982).
30. R. F. Czaja, C. B. Snowden, R. J. Casady, *J. Am. Stat. Assoc.* **81**, 411 (1986).

31. C. Brown and J. Ritchie, *Focussed Enumeration: The Development of a Method for Sampling Ethnic Minority Groups* (Policy Studies Institute and Social and Community Planning Research, London, 1981).
32. J. W. Bergsten and S. A. Pierson, *Proc. Sect. Survey Res. Methods Am. Stat. Assoc.* (1982), p. 145.
33. S. Sudman and M. G. Sirken, "The sampling of disabled scientists using networks," paper presented at the 147th Annual Meeting of the American Statistical Association, San Francisco, 17 to 20 August 1987.
34. S. Sudman, paper presented at the 39th Annual Meeting, American Society of Criminology, Montreal, 11 to 14 November 1987.
35. S. Sudman, *Soc. Networks* 7, 127 (1985).
36. ——, *ibid.* in press.
37. —— and H. Freeman, "The use of network sampling for locating the seriously ill" (Working Paper, Survey Research Laboratory, University of Illinois, Urbana, IL, 1987).
38. H. Freeman *et al.*, *Health Affairs* 6, 6 (1987).
39. J. Spaeth, "Probability samples of organization and their hierarchies" (Working Paper, Survey Research Laboratory, University of Illinois, Urbana, IL, 1987).
40. G. A. F. Seber, *The Estimation of Animal Abundance* (Griffen, London, 1973).
41. Y. M. Bishop, S. E. Fienberg, P. W. Holland, *Discrete Multivariate Analysis* (MIT Press, Cambridge, MA, 1975), chap. 6.
42. W. E. Deming and N. Keyfitz, "Theory of surveys to estimate total population," paper presented at the United Nations World Population Conference, Belgrade, Yugoslavia, 30 August to 10 September 1965.
43. R. J. Magnani *et al.*, *Evaluating Censuses of Population and Housing* (Bureau of the Census, Suitland, MD, 1985).
44. C. D. Cowan, thesis, George Washington University, Washington, DC (1984).
45. ——, W. R. Breakey, P. J. Fischer, *Proc. Sect. Surv. Res. Am. Stat. Assoc.* (1986), p. 170.
46. M. R. Frankel, *ibid.*, p. 176.

Abrupt Climate Change and Extinction Events in Earth History

THOMAS J. CROWLEY AND GERALD R. NORTH

Slowly changing boundary conditions can sometimes cause discontinuous responses in climate models and result in relatively rapid transitions between different climate states. Such terrestrially induced abrupt climate transitions could have contributed to biotic crises in earth history. Ancillary events associated with transitions could disperse unstable climate behavior over a longer but still geologically brief interval and account for the stepwise nature of some extinction events. There is a growing body of theoretical and empirical support for the concept of abrupt climate change, and a comparison of paleoclimate data with the Phanerozoic extinction record indicates that climate and biotic transitions often coincide. However, more stratigraphic information is needed to precisely assess phase relations between the two types of transitions. The climate-life comparison also suggests that, if climate change is significantly contributing to biotic turnover, ecosystems may be more sensitive to forcing during the early stages of evolution from an ice-free to a glaciated state. Our analysis suggests that a terrestrially induced climate instability is a viable mechanism for causing rapid environmental change and biotic turnover in earth history, but the relation is not so strong that other sources of variance can be excluded.

THE STUDY OF EXTINCTION EVENTS DURING EARTH HISTORY was given considerable impetus by the hypothesis of an asteroid impact at the end of the Cretaceous (*1*). Further work suggested that extinctions may also be periodic and related to cycles of comet impacts (*2*). Although these hypotheses have been challenged (*3*), extraterrestrial impacts remain a plausible possibility as a mechanism for causing environmental disruptions (*4*). However, in this article we consider whether abrupt environmental change and extinction events may also result from a discontinuous climate response to slowly varying terrestrial boundary conditions; that is, under certain conditions, instabilities in the climate system can be triggered by small changes in forcing. We believe it is appropriate to examine this mechanism more closely, because there is a growing body of theoretical and empirical support for such responses in the climate system. Furthermore, the impact-extinction correlation at other extinction boundaries, although sometimes present (*5*), is not as strong as it is for the end of the Cretaceous. These results suggest the need for other mechanisms that cause abrupt environmental change.

The hypothesis of extinctions resulting from terrestrially induced climate variations, sometimes with threshold effects, has often been discussed previously as a factor that could have contributed to biotic turnover (*6*). Most of these conjectures have been from the vantage point of a geologist (*7*). We believe it is useful to examine the problem from the climate modeling perspective. Our contribution to the problem involves several new features. There is a more complete description of climate models with abrupt transitions, with some special emphasis on conditions occurring during transition periods. We also review geological evidence indicative of abrupt climate change. Finally, we compare times of extinction with climate change.

Models of Abrupt Climate Change

Theoretical support for the hypothesis of abrupt climate change is based on climate model results that suggest the presence of multiple equilibrium climate states for a given level of forcing. Transitions between states at "critical points" can be rather sudden and can be caused by small changes in forcing. Such features have been known to exist in simple climate models for some time (*8*), and there has

T. J. Crowley is a senior scientist with the Applied Research Corporation, ARC Technologies, 305 Arguello Drive, College Station, TX 77840. G. R. North is distinguished professor of meteorology and oceanography and director, Climate System Research Program, Department of Meteorology, Texas A&M University, College Station, TX 77843.