20
Survey sampling, right or wrong?

H. Kamermans
Institute of Prehistory, Reuvensplaats 4, P.O. Box 9515, 2300 RA Leiden, THE NETHERLANDS

20.1 History

After more than 10 years of silence, the 'sampling in archaeology' debate seems to have started again. Triggered by the 1959 symposium *The Application of Quantitative Methods in Archaeology* at Burg Wartenstein in Austria (Heizer & Cook 1960), sampling in archaeology became a 'hot topic' with Binford's 1964 *American Antiquity* paper entitled 'A consideration of archaeological research design'. After this article, 'new archaeology' embraced sampling.

In the seventies, two major volumes about this subject were published: *Sampling in Archaeology* edited by James W. Mueller (1975) and *Sampling in Contemporary British Archaeology* edited by John F. Cherry, Clive Gamble and Stephen Shennan (1978). The first book gave a good overview of sampling with emphasis on regional and site-oriented sampling. Less attention was paid to sampling at the artefact level. Read (1975) and Binford (1975) both stressed that a sampling strategy must be in accordance with the target population, the cost of obtaining samples, the degree of precision, and so on. The British book covered the same topics with more emphasis on the sampling of archaeological assemblages. Plog, both in Flannery's book (Plog 1976) and in his 1978 *American Antiquity* article, compared different kinds of sampling methods for regional surveys.

During the eighties, everybody was sampling and only Hole (1980) and Nance (1983, 1990) seemed to be interested in the theoretical background.

In 1990, the debate started again. Fish and Kowalewski edited a book that criticised sampling methods for regional surveys and, during the December 1993 Theoretical Archaeology Group conference in Durham, an entire session was devoted to the subject. The topic of that session was sampling in excavations, and all of a sudden almost everybody was against sampling.

20.2 Sampling in Archaeological Surveys

The questions people tend to ask themselves when they apply sampling to regional survey are: How big must my sample size be? Should I stratify? How does the natural environment and my research strategy influence my results?

The first question, about sample size, remains a problem. There are ways to calculate sample size, but nobody uses them. Plog (1976, 148) in Mesoamerica used a 10% sample of the total area of interest and Shennan (1985, 11) in East Hampshire a 20% sample, but neither explains why these proportions were selected.

Read (1975, 59) has answered the second question, should I stratify, satisfactorily. One should only stratify if the resulting strata have greater internal homogeneity with respect to the variable of interest than does the region as a whole.

The third question, how the natural environment and the research strategy influence results, was tackled by Hodder and Malone (1984). Slightly later, Shennan (1985) wrote an interesting study about the influence of what he called field effects and walker effects.

However, little attention seems to be paid to sampling errors and the general problem of nonresponse. An example will illustrate this problem.

20.3 The Agro Pontino Survey

During the 1980s, I was co-director of the Agro Pontino Survey, a regional archaeological project in the Pontine marshes, a coastal area 80 km south of Rome (Voorrips, Loving & Kamermans 1991).

During a period of ten years we collected an archaeological, palynological and pedological data set and reconstructed the landscape from the last Ice Age until the Roman period. Our major goals were to study the change from the Middle Palaeolithic to the Upper Palaeolithic, study land use from the Middle Palaeolithic to the Bronze Age and test whether land evaluation is a valuable method for archaeology (Kamermans 1993).

The total area of the Agro Pontino is approximately 750 km^2. Since there was neither the money nor the manpower to survey the entire area, it was decided to sample it (Loving, Kamermans & Voorrips 1991). A multistage approach was used (Redman 1973) consisting of three stages: an exploratory phase, a probabilistic phase and a problem-oriented phase. In the exploratory phase data were collected for calculating the sample size for the second phase. During the probabilistic phase data were collected relating to questions about the archaeological record of the area, and the last or problem-oriented phase was simply directed towards collecting more data for specific research questions.

During the exploratory phase two kinds of survey were undertaken: spot-checking different soil units, and continuous surveys in specially selected areas. This provided the data to calculate the sample size for the probabilistic phase, which gave prior knowledge about the target (Binford 1964, 1975; Hole 1980; Read 1974).

The following formula was used to calculate the sample size for the probabilistic phase (Mendenhall *et al.* 1971, 46) with the unit of observation, the agricultural field, being the sample element:

$$n = \frac{Npq}{(N-1)D + pq}$$

where:

n = sample size required;

N = number of elements in the sampling frame;

p = proportion of interest;

q = 1- p;

$D = B^2/4$, where B is the bound on the error of estimation.

It should be noted that B is in the same unit terms as the estimators p and q. That is, $B = .05$ means an absolute error of ±0.05.

For calculating absolute numbers rather than proportions, the formula was:

$$n = \frac{N\sigma^2}{(N-1)D + \sigma^2}$$

where:

σ^2 = the population variance, estimated by the sample variance;

all other symbols are the same as above (Mendenhall *et al.* 1971, 40).

Several different research questions were asked about the area and the data collected during the first phase were used as estimators of variance. The required sample size varied between 293 and 670 fields.

A systematic unaligned transect sampling design was chosen and, after determining how many fields could be expected to occur along a transect running from the sea to the mountains, the region was subdivided into five areal blocks and the location for the transect selected using a random method. This gave 750 sampling units (more than the required 670).

The archaeology in the Agro Pontino is abundant. Almost 70% of the fields on the transects contained archaeological finds (Loving *et al.* 1991, 71). A field could contain one or more findspots (separated clusters of archaeological material) or a findspot could be in more than one field if the distribution of materials continued over field boundaries. Unfortunately, the sample did not produce enough 'dateable' findspots for the land evaluation application. It was difficult to make reliable statements about land use in the past in the Agro Pontino using simple statistical tests like the chi-squared test because the number of usable observations were too few (Kamermans 1993, 200). In other words, the sample size was not big enough to include a sufficient number of findspots with chronologically significant artefacts.

20.4 Full-coverage survey

There are a number of ways to increase sample size. First, the region can, of course, be surveyed completely. As Flannery said (1976, 1322): 'If you can survey your entire region meter by meter, do so in preference to sampling'. Kowalewski (1990) suggests a number of reasons why full-coverage information differs fundamentally from data generated by sample surveys, two of which are relevant in this discussion:

1. Full-coverage surveys generate larger data bases and more variability
2. Full coverage is recommended for the recovery of rare items

Of course it is possible, as Plog (1990, 245) points out, to create large data bases by sampling, but in general a full-coverage survey will produce more data than a sample survey.

Certainly a larger data base, more variability, and the locations of rare items would have been welcome results of the Agro Pontino survey, but even bearing in mind that a full-coverage survey is not proportionally more expensive than a sample survey, it could not have been done. The survey covered a little over 2% of the area in 7 field seasons. To cover the entire area, we would probably be out there for the rest of our lives!

20.5 Nonresponse

Even if, for whatever reason, the entire area cannot be covered, there are ways to ensure a sufficiently large sample size. The main reason that the sample size turns out to be too small is due to nonresponse (Kalton 1983; Muilwijk *et al.* 1992). Nonresponse is the failure to collect survey data from some sampled elements. This leads to a smaller sample size. This is one of the most important problems in sampling. A percentage of 70 to 80% nonresponse is no exception, and can bias results to a considerable extent. There are two types of nonresponse: total nonresponse and item nonresponse. Total nonresponse occurs when no information is collected for a sampled element. Total nonresponse in terms of an archaeological survey occurs if the selected field or area cannot be surveyed. Item nonresponse is when some but not all the information can be collected. Note that nonresponse is part of the non sampling errors (nonresponse is not a consequence of sampling) which means that these problems also occur when a total survey is undertaken.

Fortunately, there are solutions. The solutions for total nonresponse are easy. First there is revisiting (Kalton 1983, 65; Muilwijk *et al.* 1992, 244). If a field cannot be surveyed because of crop conditions or the absence of the owner, it is revisited some time later.

Another solution is to replace the nonrespondent with another element (Muilwijk *et al.* 1992, 244); in other words, to survey a different field. A problem arises if a selected field will 'never' be available for survey because, for instance, there is a house built on that spot. In such

cases, reserve elements that are not in the original sample can be surveyed, which allows the sample size to remain the same. The Agro Pontino survey used 1:5000 aerial photographs of the transect which showed 250m on both sides of the transect line proper. The field crew would select an adjacent field on the aerial photograph if the priority field, the field crossed by our transect line, could not be surveyed (Loving & Kamermans 1991, 80). In this way all the fields on both sides of the transect line served as potential reserve sample elements.

The third solution is to weight the fields again. This is a way of stratification after selection or poststratification (Kalton 1983, 74). An example of this might be that in one geographical region the nonresponse is particularly high. The sample can be weighted to a known population distribution to compensate for nonresponse. This means that some information has to be available for both respondents and nonrespondents. In this case, the stratum size of the nonrespondents can be used, among other things. Weights should be developed carefully as errors can be made easily.

Thus, in general there seems to be no problems with total nonresponse in sample surveys that would not also be encountered in full-coverage surveys.

The other form of nonresponse, item nonresponse, plays an often neglected but important role in archaeology. All kinds of information may be recorded from a surveyed field, including variables relating to the environment, to the field strategy, etc. If archaeological material is found, then functional and/or chronological aspects of the findspot are of interest. Usually, findspots of the same age will be plotted on a regional map. However, in order to do so, a field with a findspot must score on the variable 'date', which is one of the items of a respondent. If there are almost 500 findspots and only 230 of them have a score for the variable 'date' (as there were after the Agro Pontino probabilistic survey), then the item nonresponse for this variable is 54%. Of course, this is a material culture problem, but it still represents item nonresponse.

If a literature search is conducted for a possible solution for item nonresponse, the answer given is imputation (for example, Kalton 1983, 67). This means that the research variable of the nonrespondent is given the same value as that of the respondent that has the most values of the other variables identical to those of the nonrespondent. In other words, values are assigned for the missing responses using the responses in otherwise similar cases.

In this case, giving a date to a findspot in this manner this doesn't seem like the right solution. If all other variables are equal, does that mean that findspots will have the same date? Most probably this form of item nonresponse is not evenly spread across the population. For instance, findspots of some periods are easier to date than those of other periods.

20.6 Conclusions

Have archaeologists fooled themselves for the last decade by saying that good archaeological regional analysis could be done on the basis of a sample? Not for some specific research questions – if general trends are of interest, then sampling is appropriate. But if one wants to do spatial analysis or record sites and monuments, one should always try to do a full-coverage survey (Orton 1992, 138).

I have the feeling that item nonresponse is a neglected topic in regional archaeological research and I hope that an answer to this problem can be found. Even if, as in the case with the Agro Pontino Survey, the dating problem is mainly a matter of 'bad' material culture, a larger sample could have overcome this item nonresponse problem.

For the particular problems of the land evaluation approach in the Agro Pontino another solution could have been to adopt a stratified sampling design instead of the nonstratified design used. The area could have been stratified according to the landunits used for the land evaluation.

I think that we are not much further along than in 1976, when Flannery made his remarks. We should survey our entire region meter by meter in preference to sampling, but the 100% survey does not seem very realistic. In any case, we should pay more attention to the problem of nonresponse and, of course, be very cautious with our results.

Acknowledgements

I would like to thank Susan H. Loving for commenting on the text and correcting the English.

References

BINFORD, L. 1964. 'A consideration of archaeological research design', *American Antiquity*, 29, 425–441.

BINFORD, L. 1975. 'Sampling, judgement, and the archaeological record', in J. W. Mueller (ed.) *Sampling in Archaeology*, The University of Arizona Press, Tucson, 251–257.

CHERRY, J. F., GAMBLE, C. & SHENNAN, S. (eds.) 1978. *Sampling in Contemporary British Archaeology*, British Archaeological Reports British Series 50, Oxford.

FISH, S. K. & KOWALEWSKI, S. A. (eds.) 1990. *The Archaeology of Regions: A Case for Full-Coverage Survey*, Smithsonian Institution Press, Washington.

FLANNERY, K. V. 1976. 'Sampling on a regional level', in K. V. Flannery (ed.) *The Early Mesoamerican Village*, Academic Press, New York, 131–136.

FLANNERY, K. V. (ed.) 1976. *The Early Mesoamerican Village*, Academic Press, New York.

HEIZER, R. F. & COOK, S. F. (eds.) 1960. *The Application of Quantitative Methods in Archaeology*, Wenner Gren Foundation for Anthropological Research, New York.

HODDER, I. & MALONE, C. 1984. 'Intensive survey of prehistoric sites in the Stilo region, Calabria', *Proceedings of the Prehistoric Society*, 50, 121–150.

HOLE, B. L. 1980. 'Sampling in Archaeology: a critique', *Annual Review of Anthropology*, 9, 217–234.

KALTON, G. 1983 *Introduction to Survey Sampling*, SAGE Publications, Newbury Park.

KAMERMANS, H. 1993. *Archeologie en Landevaluatie in de Agro Pontino (Lazio, Italië)*, Unpublished PhD thesis, Amsterdam.

KOWALEWSKI, S. A. 1990. 'Merits of Full-Coverage Survey: Examples from the Valley of Oaxaca, Mexico', in S. K. Fish & S. A. Kowalewski (eds.) *The Archaeology of Regions: A Case for Full-Coverage Survey*, Smithsonian Institution Press, Washington, 33–85.

LOVING, S. H., KAMERMANS, H. & VOORRIPS, A. 1991. 'Randomising our walks: the Agro Pontino survey sampling design', in A. Voorrips, S. H. Loving & H. Kamermans (eds.) *The Agro Pontino Survey Project*, Studies in Prae- en Protohistorie, 6, Universiteit van Amsterdam, Amsterdam, 61–78.

MENDENHALL, W., OTT, L. & SCHEAFFER, R. L. 1971. *Elementary Survey Sampling*, Duxbury Press, Belmont, California.

MUELLER, J. W. (ed.) 1975. *Sampling in Archaeology*, The University of Arizona Press, Tucson.

MULWIJK, J., SNIJDERS, T. & MOORS, J. 1992 *Kanssteekproeven*, Stenfert Kroese Uitgevers, Leiden.

NANCE, J. D. 1983. 'Regional Sampling in Archaeological Survey: The Statistical Perspective', *Advances in Archaeological Method and Theory*, 6, 289–356.

NANCE, J. D. 1990 'Statistical sampling in archaeology', in A. VOORRIPS (ed.) *Mathematics and Information Science in Archaeology: A Flexible Framework*, Holos verlag, Bonn, 135–163.

ORTON, C. 1992. 'Quantitative methods in the 1990s', in G. Lock & J. Moffett (eds.) *Computer Applications and Quantitative Methods in Archaeology 1991*, British Archaeological Reports International Series S577, Oxford, 137–140.

PLOG, S. 1976 'Relative Efficiencies of Sampling Techniques for Archaeological Surveys', in K. V. Flannery (ed.) *The Early Mesoamerican Village*, Academic Press, New York, 136–158.

PLOG, S. 1978. 'Sampling in archaeological surveys: a critique', *American Antiquity* 43, 280–285.

PLOG, S. 1990 'Some thoughts on full-coverage surveys', in S. K. Fish & S. A. Kowalewski (eds.) *The Archaeology of Regions: A Case for Full-Coverage Survey*, Washington. Smithsonian Institution Press, 243–248.

READ, D. W. 1975. 'Regional Sampling', in J. W. Mueller (ed.) *Sampling in Archaeology*, The University of Arizona Press, Tucson, 45–60.

REDMAN, C. L. 1973. 'Multistage fieldwork and analytical techniques', *American Antiquity* 38, 61–79.

SHENNAN, S. 1985. *Experiments in the Collection and Analysis of Archaeological Survey Data: the East Hampshire Survey*, University of Sheffield, Sheffield.

SHENNAN, S. 1988. *Quantifying Archaeology*, Edinburgh University Press, Edinburgh.

VOORRIPS, A., LOVING, S. H. & KAMERMANS, H. (eds.) 1991. *The Agro Pontino Survey Project: Methods and preliminary results*, Studies in Prae- en Protohistorie, 6, Universiteit van Amsterdam, Amsterdam.