14
Conservation condition surveys at the British Museum

M. N. Leese[1] and S. M. Bradley[2]

[1]Dept. of Scientific Research, [2]Dept. of Conservation, British Museum, Great Russell Street, London, WC1B 3DG, U.K.

14.1 Introduction

14.1.1 Background: the Museum of London surveys

Conservation condition surveys can be used for auditing collections and for internal management proposes; they can also contribute to scientific studies, for example on the effect of conservation treatments or storage environment. A series of such surveys, funded by what was then the Office of Arts and Libraries (OAL), has been carried out within the last few years at the Museum of London (MoL); these have usefully focused attention on survey methodology in this context. A number of Museum of London Conservation Department internal reports are available, and a summary of the work, including information on the statistical procedures, was given at the 1991 CAA Conference (Keene & Orton 1992). Discussion at the conference suggested that statistical surveys would become increasingly common, not least because of increasing governmental pressure on museums to provide performance measures.

A Working Party overseeing the Museum of London work made various observations and recommendations which they hoped might be the basis for a quasi-standardised approach that other museums could use for their own condition surveys. The final MoL report (Keene 1991) also put forward ideas for future research, suggesting that surveyor bias, the differences between collections and testing of condition changes over time should be investigated. It also suggested that blind testing of conservator assessments should be carried out within institutions who were not members of the Working Party (such as the British Museum).

As far as the authors are aware, the work reported in this paper has gone the furthest towards addressing these suggestions. In addition, the work at the British Museum (BM) has looked at computer data-gathering techniques using a palm-top Psion Series 3 computer. It was also decided to test the practicality of the MoL sampling methods for surveying archaeological collections in the BM. This paper describes the additions and modifications which were made to the survey methodology suggested by the MoL in order to meet the needs of the BM and pointers for conduct of future surveys, especially of archaeological material, are given on the basis of experiences gained so far.

14.1.2 British Museum archaeological iron surveys

A series of surveys of the condition of archaeological iron in the BM has been planned. Three aims of the surveys are related to conservation practices. The first of these aims is to identify whether iron objects which have been stripped are more or less stable than objects which have not been stripped; the second is to establish the effects of stabilising conservation treatment; the third is to establish the effects of the environment in the store. It was also intended that the conservation requirements of objects in particular need of conservation treatment should be identified during the course of the surveys.

To pursue these aims, it was decided to characterise each of the discrete subcollections of archaeological iron, partly to assess their overall condition in order to identify any gross differences attributable to environmental or storage conditions, and partly to identify individual objects which could be studied more closely or monitored over time. This paper is concerned with the methodology for assessing overall condition, taking into account the conservation aims and the particular nature of archaeological collections.

The examples described here are taken from the first subcollection considered, the Anglo-Saxon iron stored in an area known as the Sturge Basement, referred to in this paper as the 'Sturge survey'. The 3000 or so objects in this subcollection are Anglo-Saxon iron objects such as knives and spears, some newly excavated, some previously conserved, stored mainly in boxes and on shelves. The boxes are nearly all of a similar shape, with the objects arranged in an apparently haphazard fashion within them, whereas on shelves the objects tend to be laid out in a more ordered fashion.

14.1.3 Definitions of terms

The key variable, a concept introduced by Keene and Orton (1992), is that variable (for example, the percentage requiring conservation) the estimation of which determines the sampling design. Two-stage sampling involves a first stage of sampling, in this case the storage units, and then a second stage, in this case the objects within the storage units. The units can be viewed as strata, i.e. subpopulations which are sampled separately, although the term stratum would be more likely to be applied to a group of units having something in common, for example all the storage units relating to a given excavation site. The storage units can also be termed clusters, since they consist of groups of elements (objects) which are physically contiguous and could, for convenience, be sampled together. In single-stage cluster sampling there is no subsampling within clusters, data on all objects being observed, so that the unit is characterised completely. This is sometimes known as cluster sampling for short. These terms, and others used in the following, are standard and are defined in text books on statistical sampling, for example Cochran (1963).

A condition survey is designed to characterise a complete collection, usually on the basis of a statistical sample, and leads to a statement such as 'an estimated 90% of the collection requires conservation'. For a pure condition survey, the MoL suggests the use of four condition grades, designated C1–C4:

- C1 *Good*. Object in the context of its collection is in good conservation collection or is stable
- C2 *Fair*. Fair condition, disfigured or damaged but stable; needs no immediate action.
- C3 *Poor*. Poor condition and/or restricted use and/or probably unstable, action desirable.
- C4 *Unacceptable*: Completely unacceptable condition and/or severely weakened, and/or highly unstable and actively deteriorating and/or affecting other objects; immediate action should be taken.

Conservation surveys, on the other hand, are carried out with a view to identifying the conservation status of all the objects in the collection. Four categories of treatment A–D are defined as follows:

- A No work needed
- B Low conservation priority (In stable condition, but some work desirable when other priorities and/or resources permit).
- C Medium conservation priority (Not in immediate danger, but needs essential work).
- D High conservation priority (e.g. active deterioration).

It can be seen that the definitions of condition grades and conservation categories contain similar elements. For instance if an object is designated category D in a conservation survey, it would also be designated condition C4 in a condition survey. Nevertheless the definitions are subtly different, and both involve some degree of assessment of priorities and resources external to the object being considered. This is discussed in more detail in this paper.

14.1.4 Inter-observer agreement

Until recently little was known about the extent of inter-observer agreement on conservation assessments. Prior to starting the Sturge survey, therefore, an experiment was performed in which 50 iron objects were assessed by 22 observers, mainly conservators, from three institutions. The participants were asked to assess the condition of each object as A, B, C or D, without previous sight of the objects and without conferring. The full results have been discussed by Newey *et al* (1993), but two essential points emerged: that staff of different institutions differed significantly on average in their view of the percentage of objects requiring conservation, and that variation between individuals was substantial even within institutions. For example, within the BM, the percentage of objects estimated by different individuals as requiring conservation ranged between c. 50% and c. 80%. The typical within-institution standard deviation (averaged over the three institutions) was ± 10%. The implication for sampling design is that there is little point in aiming for high sampling precision through large sample sizes, unless this standard deviation can be reduced.

Discussion of the results after the experiment revealed that one possible source of disagreement between observers, particularly between those in different institutions with different management and training practices, was the extent to which the curatorial priority or 'value' of the object should enter into the assessment. For most condition surveys the value would not be relevant. However, as noted above, the definitions commonly in use are slightly ambiguous about this point. Also, at the BM surveys tend to have a dual purpose as both condition and conservation surveys and conservators will naturally have in mind the value of the object when making their assessment.

14.1.5 Redefinition of conservation need

A possible way round the problem of ambiguous definition is to use a double scale covering both conservation condition and curatorial value; these can be combined to give a single assessment of conservation priority. One could envisage several methods for combining scores on condition and value: the table below shows the result of one of these, based on assigning scores of 0–3 to conditions A-D. Given value scores of 0–3, one obtains scores 0–9 by multiplication (Table 14.1).

The numerical values in the body of the table have no meaning, other than to rank the possible combinations and hence provide priority categories. Thus if four conservation priority categories were required, priority I could be assigned to score 0, priority II to scores 1–2, priority III to 3–4, and IV to scores 6 and 9.

Suppose the curatorial value and conservation conditions for a group of objects were both equally divided among their respective four categories so that any particular object was equally likely to appear in any one of the 16 possible combinations of condition and value. The net result of this division would be that roughly 44% ($^7/_{16}$) of objects would be assigned to conservation priority I and 19% ($^3/_{16}$) each to priorities II, III and IV. With a more realistic division of conservation categories into A:60%, B:28%, C:8% and D:4% (as found in the Sturge survey), again with equally divided curatorial values, the distribution of conservation priorities would be I:70%, II:16%, III:10% and IV:4%.

Keeping the two types of assessment separate in the first instance and combining them formally is recommended, although it is recognised that a potential problem with this approach might be that conservators may not feel confident in assigning a definite curatorial value. It should also be pointed out that the suggestion above is only one of many possibilities and that it is primarily a conservation management problem to

Curatorial value	*Conservation Condition (A best, D worst)* A(0)	B(1)	C(2)	D(3)
3 (high)	0	3	6	9
2	0	2	4	6
1	0	1	2	3
0	0	0	0	0

Table 14.1: Scoring conservation priority.

determine the link between conservation condition and curatorial value. What the above analysis demonstrates is that even given *a priori* evenly spread curatorial values, the final spread of conservation priority categories may not be the same as the spread of conservation condition categories.

14.2 The Sturge Survey

14.2.1 Data collection methods

In the normal course of a conservation survey, a standard paper form is used. The fields used vary with the type of object being surveyed. For instance the deterioration processes observed on objects made of iron are different to those on objects made of plastic. The data recorded are the registration number of the object, its type, the type of damage (if applicable), storage conditions, the number of hours of conservation work needed, as well as the conservation condition category; the latter was the key variable for the Sturge survey.

In a pure condition survey, such detailed information is not necessary. The minimum information required must of course include the key variable(s), but it must also include the data used to weight observations appropriately in the final statistical estimates. Exactly what is recorded depends on the sampling procedure adopted. For the Sturge survey, whose sampling procedure is described below, the weights were based on the numbers of objects in the storage unit (box or shelf) and the number sampled from the unit.

A 0/1 'flag' to indicate whether or not any particular object had been sampled at random is also necessary if conservators wish to note objects in bad condition as part of a conservation survey, independently of the statistical sampling procedure adopted for the condition survey. It is important to alert conservators to the need for such a flag when entering data to avoid inadvertent mixing of random with non-random samples.

A palm-top computer (a Psion 3 Organiser) was used to collect data without the use of paper, storing the data on 'flashpacks' (removable solid state disks). The survey took place over about 6 months and, approximately weekly, the data were transferred via a serial link to a dBase IV database on a personal computer (PC). This proved to be a quick and convenient way of data capture, but certain points should be considered when planning palm-top data collection:

- The PC file must be of exactly the same format as the palm-top file, with fields long enough for the longest information (short codes should of course be used where possible.)
- Fields which are likely to contain the same data for several contiguous records (e.g. the name of the conservator) should be placed together in a block, preferably at the start of the list of fields, so that data can be copied from one record to the next automatically.
- If no automatic data validation has been built into the system (possible but time-consuming), it is advisable to transfer data to a 'hold' file on the PC before final transfer, in order to check that the data have been correctly formatted.
- The flashpacks can act as temporary backup while the survey is proceeding, with hard-copy produced whenever the data are transferred to the PC; at the end of the survey a backup of the PC file should be made.

If these points are observed palm-top computer entry can be relatively efficient and error-free.

14.2.2 Sampling procedures used in the Sturge survey

The sampling criteria for the Sturge survey differed in two main respects from those underlying the Museum of London schemes. Firstly, it was decided that every storage unit should be opened (to avoid missing important objects in need of conservation) and secondly the overall cost of the survey was not fixed; rather, the survey was to be continued until a satisfactory result had been obtained. In view of the inter-observer variation discussed earlier, which is of course not reduced by increased sampling, 'a satisfactory result' was taken to mean achieving a sampling precision (standard error) of less than ± 10% in the estimate of the percentage of objects requiring conservation.

The sampling scheme adopted was to open every box and then to sample one object from each, but a few boxes and all of the shelves were sampled more intensively. This was 'overkill' from a statistical point of view (c. 600

objects sampled out of c. 3000) but was necessary to meet the conservation needs of the survey (to view every box). The boxes and shelves were treated as strata and the results were weighted by the numbers of objects in the stratum. Standard errors were approximately estimated on the basis of simple random sampling. Use of the technique of 'collapsed' strata, and consideration of the data from the strata which had been sampled more intensively, showed this to be a reasonable approximation since the variation among objects within strata was similar to that over the whole data-set.

The main problem with this approach lay with the practical procedures for random sampling within the boxes. While the 'route-following' method worked well for objects laid out on shelves (see Figure 14.1a), it was difficult and time-consuming to apply to objects laid out in an apparently haphazard fashion in boxes; thus for boxes a grid method was used (see Figure 14.1b). The latter, while satisfactory for proportions, tends to overestimate the hours of conservation work required, because larger objects have a greater chance of being chosen. It was therefore decided to rethink the approach for the next survey, and to use a method which would avoid sampling within storage units and provide satisfactory estimates of standard errors, and yet would still meet conservation needs by allowing access to all storage units.

14.3 Future Sampling Approaches

14.3.1 Single-stage cluster sampling

A feasible solution to the sampling problem discussed above is to use single-stage cluster sampling, in which storage units are sampled and then data for all objects within those storage units collected. This has the advantage that no choices need to be made about which objects to sample. However to achieve the same precision as a simple random sample (or indeed a two-stage sample), cluster sampling generally requires a larger sample size, as measured by the design effect, defined as the ratio of variance of the estimate of the key variable to the equivalent variance for a simple random sample of the same size. The effect depends on the degree of homogeneity (in terms of the key variable) within units and numbers of objects in the units, and it determines the sample size relative to that required for a simple random sample having the same precision (see Moser and Kalton 1971). As an example, assuming a correlation of 0.2 between the key variable for pairs of objects in a given unit, and an average number of objects per unit of 6 (typical values in this context), the design effect would be 2, implying that one would require to sample approximately double the number of objects compared to a simple random sample to obtain same precision.

The formulae for the estimates and their (squared) standard errors for variables such as total hours of work and also for proportions, are given in the appendix (below); they can be seen to be similar to those for two-stage sampling (see Keene & Orton 1992), except that the second term in the variance formula is zero because the second-stage sampling is a complete survey for the unit and is therefore error-free. The formula for the design effect is also given.

On completion of the next in the envisaged series of archaeological iron surveys, it will be possible, using the final estimates obtained and the relative timings involved in sampling and observing objects and the overheads of opening boxes, to compare the design effects for cluster sampling and for the two-stage scheme proposed by the Museum of London team. It should also be possible to consider at that stage the relative cost-effectiveness of the two schemes, bearing in mind that because of conservation requirements of the BM, all storage units will probably be opened, although not necessarily sampled in detail, whichever scheme is adopted.

14.3.2 Summary of recommended approach

Given that in BM surveys it will usually be the case that all storage units will be opened, and given the conclusions regarding sampling within boxes in the Sturge survey, the approach recommended for similar types of collections, termed multiphase stratified cluster sampling, is summarised below.

The objects should be stratified to keep similar material together (e.g. newly excavated objects), since the characteristics of objects may vary between strata. If equal probability sampling of units is adopted, the sample will be self-weighting and a simple average or sum of the estimates from different strata will give unbiased estimates for the whole subcollection, with standard errors combined in the usual simple manner for means or sums. (This would not be the case for other choice of within-stratum sampling). Thus if newly excavated material is stored in 20 units and other objects in 10, then the cluster-sampled units would sensibly be divided between strata in the same ratio, i.e. 2:1.

If every storage unit is to be opened, for example for conservation needs, the total numbers of objects and perhaps other limited data can be collected while doing so; this can be considered a first phase of data collection, while the cluster sampling of a proportion of units for more detailed information can be seen as a separate exercise. The term multiphase, which describes this approach, does not have any timing implications, but rather that more than one level of information is being sought. One by-product of this approach is that the total numbers of objects will be known, rather than being estimated as in a two-stage survey where only a sample of units is opened.

Cluster sampling implies the collection of data on all objects in those units that have been sampled. This is recommended primarily because of the difficulty of enumerating, and hence sampling from, haphazardly arranged objects. For orderly objects, one might prefer to sub-sample within units, as the Museum of London team recommend; other things being equal, this would lead to better precision for the same sample size.

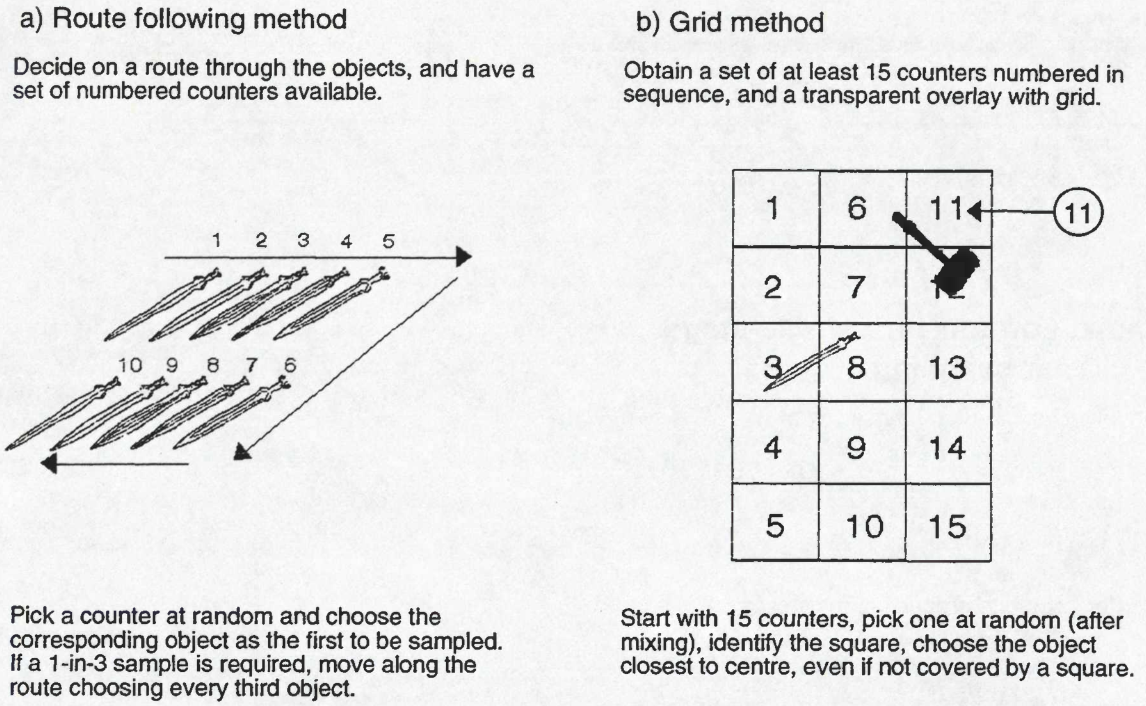

Figure 14.1: Two types of sampling. a) 'Route-following' method where objects are counted along an arbitrary but predetermined route. For a sample of 1 in n objects, every nth object is chosen, starting at a random number: strictly a systematic sample but will give an pseudo-random selection in most cases. b) Grid method, easier for haphazard arrangements of objects but only satisfactory for percentages in different condition, not total hours of work since this is affected by the size of the object.

14.4 Conclusions and Future Work

The use of computerised data collection has been successful but needs careful planning of the record structure; care must be taken to include the statistical weighting information and to make backups and hard-copy records as the survey proceeds.

Further consideration should be given to definitions of condition and it is suggested that the concepts of the condition and the curatorial value of the object should be separated in the first instance, to be combined if necessary to give conservation priority.

The sampling methodology recommended by the OAL Working Part on the basis of the Museum of London surveys may need to be adapted if surveys are to be both condition and conservation surveys as at the BM and/or where objects are haphazardly arranged within storage units. A slightly different alternative sampling scheme (multiphase stratified single-stage cluster sampling) has been proposed which copes with these situations.

Although the overall cost of the survey is unlikely to be the overall determining factor in the design of a survey at the BM, it will be important to assess the cost effectiveness of any proposed sampling scheme and investigating this will be a feature of the next survey to be performed. The concept of the 'design effect' will be of use here.

Future work will continue with

1. further cross-sectional surveys of further subcollections of archaeological iron, using the experiences gained from the Sturge, and
2. longitudinal studies of certain individual objects in the Sturge basement.

The latter will be aimed at monitoring the condition of objects over time and it is likely that this will involve stratification by existing condition, type of object and degree of previous conservation treatment.

Acknowledgements

The survey in the Sturge Basement was performed by Fleur Shearman and Ruth Goldstraw, of the British Museum's Department of Conservation. The authors are grateful for their comments on the survey methodology. A report on the results of the Sturge survey is being prepared.

References

COCHRAN, W. G. 1963. *Sampling Techniques*, (2nd Ed.), Wiley, New York.

KEENE, S. 1991. *Collections Condition Surveys*, Museum of London Conservation Department Research Reports, Museum of London, London.

KEENE, S. & ORTON, C. 1992. 'Measuring the condition of museum collections' in G. Lock & J. Moffett (eds), *Computer Applications and Quantitative methods in Archaeology 1991*, International Series S577, 163-166, British Archaeological Reports, Oxford.

NEWEY, H., BRADLEY, S. M. & Leese, M. N. 1993. 'Assessing the condition of archaeological iron: an intercomparison', in Bridgland, J. (ed), *Tenth Triennial Meeting of ICOM Committee for Conservation*, Washington, DC Preprints Vol 2, 786-79.1 International Council Museums Committee Conservation, Paris.

MOSER, C. A. & Kalton, G. 1971. *Survey Methods in Social Investigation*, Gower, Aldershot.

Appendix: Formulae for single-stage cluster sampling

i) Continuous (Y): e.g. total hours of work:

$$\hat{Y}_R = M_o \sum y_i / \sum M_i$$

$$V(\hat{Y}_R) = (1-f) N^2 / n \left[\sum M_i^2 (\bar{y}_i - \bar{y})^2 / (N-1) \right]$$

ii) Proportions (p): e.g. proportion in category A:

$$\hat{p} = \sum a_i / \sum M_i$$

$$V(\hat{p}) = (1-f)/n \sum M_i^2 (\bar{p}_i - \bar{p})^2 / \left[\overline{M}^2 (N-1) \right]$$

M_o total number of objects in the collection

M_i number of objects in the ith cluster (storage unit)

$\overline{M}$ mean number of objects per cluster

N total number of clusters

n number of clusters sampled

f proportion of clusters sampled (sampling fraction)

y_i total (Y) in ith cluster (eg hours of work)

$\bar{y}_i$ mean (Y) per object in ith cluster

$\bar{y}$ overall mean (Y) per object sampled

$\hat{Y}_R$ ratio estimate of Y for population

a_i number of objects in category (eg in A) in ith cluster

p_i proportion in ith cluster

$\hat{p}$ overall proportion in sampled clusters

V variance (squared standard error)

Approximate design effect for cluster sampling (ratio of variance to variance with simple random sampling):

$$1 + (\overline{M} - 1)r$$

where r is the correlation coefficient between individuals within clusters (of Y or p).