The Use Of Condition Score To Determine Glycerol Concentration In The Treatment Of Waterlogged Archaeological Leather. An Empirical Solution.

Kirsten Suenson-Taylor **and** **Dean Sully**

Museum of London Archaeology Service,
Walker House, 87 Queen Victoria Street,
London, EC4 4AB
UK

Museum Of London,
150 London Wall,
London, EC2Y 5HN
UK.

Introduction

In recent years the treatment of waterlogged archaeological leather, unlike wood, has received little attention in the conservation literature. The selection of treatments for wet leather is generally thought to be less critical than that for wet wood. However variations in the treatment of leather do affect the success of the treatment process. By examining these effects we may identify which are the critical elements of the treatment.

At the Museum of London (MoL) the treatment of waterlogged leather with glycerol impregnation and freeze drying, follows a method developed in the early 1980's for the efficient treatment of large quantities of material (1). This study aims to build upon that work, by establishing the most effective concentration of glycerol to use with each individual artefact. To do this it has been necessary to implement a procedure to evaluate the relative success of treatments. This has been carried out using an experimental design similar to those commonly used in industrial and medical research (2).

This study forms part of a larger research project reviewing our approach to the conservation of archaeological leather. It is hoped that other aspects of the leather treatment process will be investigated by the authors over the next few years.

Leather treatment

Leather is treated by immersion in a glycerol solution followed by freeze drying. The concentrations of glycerol previously used at the MoL have been 15%, 20% or 25%. The use of different concentrations reflects concerns about the appearance of the material after conservation, rather than any adaptation of the treatment to the condition of individual artefacts. Concentrations ranging from 10-50 % have been recommended (3) and it has been suggested that lower shrinkage values are achieved by using higher glycerol concentrations (4).

The treatment of wet leather, carried out at the fieldwork stage of archaeological projects, is designed to stabilise material for inclusion in the site archive (5), according to the requirements of receiving body (6). It is inappropriate to attempt to return the artefact to its perceived condition prior to burial. It is however necessary to maintain the physical properties of the wet material in the treated artefact. The material, once treated is expected to be stable in long term storage.

The Use Of Glycerol

Leather is found in abundance in the wet anoxic conditions of the Thames waterfront sites. Although the materials introduced during manufacture and use may have been lost during burial, the leather fibre network can survive, with degrees of de-polymerisation and de stabilisation. Deterioration may occur down to the level of the collagen macromolecule (7). In waterlogged burial conditions, water molecules may temporarily satisfy polar areas on the collagen molecule and physically hold fibres apart.

Glycerol (1,2,3,propanetriol) is a relatively small molecule with two primary and one secondary hydroxyl groups. The size of the glycerol molecule, it has a molecular weight of 76-92, compared to 380-420 for PEG 400 (8), may be a contributory factor to its success in the treatment of waterlogged leather. Multiple hydrogen bonding between the three hydroxyl groups of the glycerol molecule and polar areas of the collagen triple helix, stabilises the leather fibre structure, enabling water to be safely removed (9). The molecular structure of glycerol, that is the position of polar groups and chain length, may facilitate the stabilising effect (10).

The primary aim of this study is to test whether the use of different glycerol concentrations, selected according to the condition of the leather, might improve the condition of the treated artefact. To do this, an effective method of assessing condition before and after treatment was required. It was important that this information was collected without a significant increase in the time taken for the treatment of the artefacts.

Condition Assessment Of Leather

Many factors such as dimensional stability, colour change, weight change, texture, etc. could be used to define the condition of leather. More specific quantitative characterisation of the stability of leather is problematic. Mills Reid and MacLeod (3) list the type of analytical information which might be used, identifying aspects of the leather and its burial environment. Advances have been made in characterising the condition of historical leather, however these techniques may be limited when applied to archaeological material (11). For example, microscopic analyses of hydrothermal stability using Differential Scanning Calorimetry would require multiple sampling locations in order to resolve the variation in condition across one artefact. Micro hot table methods measuring shrinkage temperature, are complicated by the presence of burial contaminants and glycerol in treated archaeological leather.

In the past at the MoL, condition information was recorded in conservation documentation by descriptive terms e.g. wet, weak, de-laminating, etc.. Such information can only be used as a rough guide to condition. A conservator describing an artefact as "good" or "bad" is putting the condition of the artefact on a scale within a personal value system. It is a straight forward process to standardise this relative scale and derive a ranked numerical system.

Pre-burial damage	**1**	**2**	**3**	**4**

Assess the amount of use wear damage in relation to object type.	Wear damage extensive over whole area	Wear damage over greater part	Isolated areas of damage not extensive.	Object intact, no wear damage
Cohesivity Consider the integrity of the object as a whole. Look at vulnerable areas liable to loss. Bear in mind nature and shape of object.	**1** Many fragments readily detached during handling.	2 Several fragments readily detached during handling.	**3** Minor areas of vulnerable fragments	4 Leather intact, no vulnerable fragments.
Friability Assess condition of fibre network and grain surface. Where grain surface is no longer present, define condition of the remaining surface.	**1** Fibres easily detached during handling, resulting in total loss of surface.	2 Greater part of surface and exposed edges liable to fibre loss.	**3** Few areas of surface liable to loss of fibres.	**4** Surface intact, no loss of fibres.
Flexibility	**1**		**2**	
Flexibility must be appropriate to the object, if flexible not so weak as to be damaging to the object. If inflexible not so brittle as to allow damage to occur during handling	Unacceptable. Weak or stiff and brittle		Acceptable, appropriate flexibility.	

Figure 1 Criterion anchored rating scale for waterlogged leather condition assessment.

The use of condition score.

A standardised method of assessment, using a criterion anchored rating scale was formulated (Fig 1). This guides the conservator to consider specific elements of an artefact's condition. Each point on the scale is specifically described, this reduces differences in interpretation of the scale and therefore limits the subjectivity of the data (12).

Four "condition score" criteria specific to leather were defined: pre-burial damage, cohesivity, friability and flexibility. Each artefact was scored 1- 4 for pre-burial damage, cohesivity and friability: flexibility was defined as unacceptable or acceptable, 1-2. It proved difficult to use flexibility as a factor to order condition, as flexibility needs to be judged in relation to the function of the artefact. In this study pre-burial damage, friability and cohesivity were combined to form the condition score.

In recording value judgements such as condition score, there will always be a degree of subjectivity. This was limited by adhering closely to the condition definitions in the criterion anchored rating scale, maintaining a consistent assessment protocol, and by using a large sample. The condition score assessed for each artefact characterises all the relevant aspects of its condition. It must therefore incorporate influencing factors such as, period, type of artefact, burial environment, the effect of pre-treatment storage, etc. The allocation of numbers to the descriptive terms provided data measured on an ordinal scale which could be tested by appropriate statistical methods. In statistical terms measurements can be made on three main scales, nominal, ordinal or interval. A nominal scale is one in which numbers are merely labels with no usable value. An ordinal scale provides numbers which are ranked values, and can therefore be ordered. A more precise scale is an interval scale, where the hierarchy of the values is precisely defined.

A treatment sheet (Fig 2) is used to record: administrative and artefact details, pre-treatment storage, the assessed condition before and after treatment, and the glycerol concentration assigned to the artefact. The information describing condition used in this study was recorded during the pre-treatment cleaning stage and after treatment, during the packing of the artefacts. The use of these forms did not increase the amount of time spent on the treatment of each artefact.

The use of clinical study procedures

Conservation treatment research questions are often similar to those of clinical treatment trials used in medical research. Previous conservation studies have used the analogy with medical trials to study long term performance of iron treatments using mortality rates (13). Treatment variation, the prescription and dosage of drugs, can be tested by implementing an intervention study, in which the effects of treatment are assessed on real patients. Data is collected on a standard form, as multiple choice answers to set questions or by recording measurements using a calibrated instrument. These procedures minimise observer bias and increase the reliability of the data (14).

The adaptation of these techniques to conservation treatment trials, help to ensure the validity of information about treatments in real situations.

Figure 2. Waterlogged Leather Treatment Sheet

Method
Information was collected to answer the following research questions:

- Does varying the glycerol concentration have a significant effect on the success of the treatment?

- Can pre-treatment condition score be used to allocate glycerol concentrations in order to improve the success of the treatment?

- Can the archaeological period of an artefact be used to allocate glycerol concentration in order to improve the success of the treatment?

The concentration of glycerol to be used in the treatment was allocated to each artefact on the basis of three different selection methods. This resulted in three sub-populations which can be studied both individually and in comparison to each other.

Group 1. In which concentration of glycerol is allocated according to the pre-treatment condition score of the artefact (condition allocated). The range of assessed condition scores was divided arbitrarily into three categories. Material in good condition (pre-treatment condition score 9-12), was treated by immersion in 15% glycerol prior to freeze drying; mid range material (pre-treatment condition score 6-9) was treated with 20% glycerol; and the poorest condition material (pre-treatment condition score 3-6) with 25% glycerol.

This population tests the hypotheses that the more deteriorated the material, the more glycerol is required to stabilise the leather structure. Further it will indicate if condition score could be used to successfully allocate glycerol concentration.

Group 2. In which concentration of glycerol is allocated according to the archaeological period of the artefact (period allocated). In previous survey work, looking at the stability of treated leather artefacts in long term storage, it appeared that Roman material was in better condition when it had been treated with lower concentrations of glycerol (15). The use of a low concentration (15%) of glycerol for the Roman material, which in general is in poorer condition counters the hypotheses that stabilising more deteriorated material requires higher concentrations of glycerol. To verify that improved, found in the surveyed material, condition was not a result of the pre-treatment condition of the leather; Roman material (first to fourth century AD) was treated with 15% glycerol; Medieval (fifth to fifteenth century AD) material was treated with 25% glycerol. Any material from an undated context was assigned a glycerol concentration in a random manner.

Group 3. In which glycerol concentration was allocated randomly. Each artefact was allocated either 15%, 20% or 25% glycerol in sequence during the treatment process. This produces a control group against which to judge the other two populations.

For each artefact, condition scores before and after treatment were compared to produce a "percentage treatment score". This figure represents the change in condition score, expressed as

a percentage of the potential to increase or decrease during treatment. The formula for calculation of percentage treatment score depends on whether condition score has increased or decreased during treatment:

1. condition score increased

$$\text{percentage treatment score} = \left(\frac{\text{(post-treatment score - pre-treatment score)}100}{\text{(highest score possible - pre-treatment score)}} \right) + 100$$

$$\text{i.e.} = \left(\frac{(X-Y)100}{(12-Y)} \right) + 100$$

X= post-treatment condition score

Y= pre-treatment condition score

2. condition score decreased

$$\text{percentage treatment score} = 100 - \left(\frac{\text{(pre-treatment score - post-treatment score) } 100}{\text{(pre-treatment score - lowest score possible)}} \right)$$

$$\text{i.e.} = 100 - \left(\frac{(Y-X)100}{(Y-3)} \right)$$

Percentage treatment score, ranging from 0 to 200%, provided a means of assessment which took into account the pre-treatment condition of the artefact and potential degree of improvement / deterioration. In this study any score of less than 100% is a poor result, since in such cases the object's condition has not improved during treatment. A score of less than 100% however, does not necessarily represent an unstable artefact. The calculation of percentage treatment score allows us to assess the change in condition of the object as a result of the type of treatment selected.

Results

The percentage treatment scores were analysed in conjunction with other object parameters for trends within the data. Table 1 presents the mean values for categories from each of the three groups.

Categories	**Sub- populations**		
	Group 1. **Condition allocated**	**Group 2** **Period allocated**	**Group 3** **Randomly allocated**
	Mean % treatment score	Mean % treatment score	Mean % treatment score
Entire population	112	96	106
Period			
Roman	107	94	101

medieval	116	93	109
Unknown	125	104	90
Glycerol concentration			
15%	110	93	103
20%	113	-	105
25%	111	104	108
Total number of cases	786	124	145

Table 1. Mean percentage treatment score for groups of data from the three sub-populations.

From these figures it is apparent that there are differences in the results from each of the three groups. These differences represent the effects of the three methods of allocating glycerol concentration. Condition allocated material tends to produce higher mean percentage treatment scores when compared with equivalent categories of material from the other two groups. The use of statistical tests provides a tool to decide whether an observed difference between two sets of data is significant. Statistical tests are divided into two major classes, parametric and non-parametric, each relies on different assumptions about the data.

The type of measurement and data produced in this study is best analysed using the more general assumptions of non-parametric statistical tests. Thus a Wilcoxon Matched Pairs Signed Ranks Test was used to compare the results from each of the three methods of treatment allocation. This test identifies the magnitude of difference between groups which contain matched data sets. This is a useful alternative to the parametric t- test, since it does not require a normally distributed population nor data measured on an interval scale (16).

Tests comparing each of the treatment selection methods verify that the differences between the data from the three groups are significant at the 1% probability level. This indicates that there are real differences in the results obtained from the different selection methods. In this study, condition allocated material results in more successful treatments than allocating treatment randomly, or by period. Period allocation is the least successful selection method. This suggests that varying procedures on the basis of condition assessment produces the most successful and consistent results.

One can also identify trends within the groups. Figure 3 shows the performance of the condition allocated material. separated according to archaeological period and by glycerol concentration. It is apparent that Roman material tends to have a lower mean percentage treatment score than material from other periods, regardless of glycerol concentration used. This suggests that this material does not respond as well to the treatment. It is possible that concentrations of glycerol outside the range used in this study may increase the efficiency of the treatment. It is also possible that parameters used in condition assessments need to be refined for Roman material.

Figure 3. Graph of condition allocated material; mean percentage treatment scores for different periods and glycerol concentrations.

The inference that period cannot be used to allocate glycerol concentration is confirmed when looking in more detail at the data for the three groups. For condition allocated material, only Roman artefacts in good condition are treated with 15% glycerol, while for period allocated data, all material regardless of condition are allocated 15% glycerol. Randomly allocated material is treated with both 15 and 25% glycerol.

Population	**Period**	**Mean percentage**	**treatment score**
		15% glycerol	*25% glycerol*
Allocated by period	Roman	94	-
Allocated by condition	Roman	110	109
Randomly allocated	Roman	101	106

Table 2. Comparison of the treatment selection processes used for Roman material

A Wilcoxon Matched Pairs Signed Ranks Test indicates that significant differences occur between the three groups as a result of treatment selection, rather than from the different glycerol concentrations. If we were to treat all Roman material with 25% glycerol, it would be no more successful than using 15% glycerol, since there is a wide range of conditions represented within this population. It is allocating concentration according to condition, rather than the period of the

artefact being treated, which appears to determine the success of the treatment. This is confirmed by examining the medieval material from each of the three groups. Again in the condition allocated population, medieval artefacts in poor condition are treated with 25% glycerol and those in good condition with 15%. In period allocated material, early medieval (Saxon) material is treated with 15%, later medieval with 25% glycerol. Once again statistically significant differences between the selection processes indicate the advantage of using condition assessment to allocate treatment.

Population	**Period**	**Mean percentage treatment score**	
		15% glycerol	*25% glycerol*
Allocated by period	medieval	96	91
Allocated by condition	medieval	118	113
Randomly allocated	medieval	104	103

Table 3. Comparison of the treatment selection processes for medieval material

Discussion

Use of clinical studies methods allows a direct assessment of real treatment situations. This approach may be more appropriate to the variability of organic archaeological materials and their treatments, than traditional analytical studies. Analytical models attempt to simplify complex systems by limiting the inherent variation, with the risk that the results from such models will bear little relation to real conservation treatments. There is large amount of variation endemic in the investigation of real treatment situations. To separate the effects of interest from the uncontrolled variation, requires the imposition of a suitable experimental design. Unless a sensible design is employed, it may be difficult to obtain valid conclusions from the resulting data. In the course of this study we have identified limitations in the design which preclude the use of more detailed comparative statistics on the data.

The main limitation derives from the selection procedure for each of the three groups. The three methods of selecting treatments were not carried out simultaneously throughout the period of study. The result is that condition allocated, period allocated and randomly allocated groups, could comprise material in significantly different condition. It is also possible that comparisons between these groups are influenced by bias introduced by different conservators carrying out the assessments. These factors may include systematic variation which limits the subtlety of information that can be derived from the data.

The broad trends in the data can clearly be identified:

- Adapting the concentration of glycerol to the condition of the object improves the success of treatments.
- The assessment of pre-treatment condition score is an effective means of defining condition and allocating treatment.
- Material identified as being in poor condition requires higher concentrations of glycerol than material in better condition.

Future work

The next stage in our project will use a randomised application of treatments to all the objects in the study. This will help to ensure that the only systematic source of variation would result from differences in the data groups. In addition it should help to ensure that any errors in the assessment procedure are random and therefore will not influence the outcome. This experimental design can be applied to the standard treatment procedure of wet archaeological leather.

In the trial, condition score will be assessed for each leather artefact and treatment assigned randomly. For instance, the first item treated is assigned 15% glycerol, the second 20%, and the third 25% (a greater range of concentrations could also be considered). The option of conducting the experiment by selecting exactly matched sets of artefacts was considered; however this is not practical when conducting such a study in normal treatment situations. If sufficient numbers of artefacts are recorded, it may be possible to select matched data sets on which to carry out statistical assessments.
The random allocation of glycerol concentration will enable us to define the effectiveness of each concentration on material from a range of conditions. This removes the arbitrary nature of assigning concentration by condition score. Analysis of the data can then identify a more sensitive allocation of treatment which is directly related to condition.

Assessment of each object must be accurate and carried out without bias. Therefore the criterion anchored rating scale will be refined. In addition, all post treatment assessment will be done in a "blind" manner, that is without knowledge of pre-treatment score or concentration of glycerol used. The accuracy of measurements derived from the criterion anchored rating scale could be calibrated by analytical procedures. Highly accurate techniques of measurement are often the most difficult to apply in treatment situations. An instrument such as a criterion anchored rating scale may be more appropriate to the large number of observations produced during treatment trials.

The use of a suitable experimental design enables the application of more sensitive comparative statistical tests. This facilitates the examination of subtle variations in the treatment. A powerful technique to examine the influence of different factors in the experimental process is Analysis of Variance. This technique is more appropriate than multiple comparison tests, such as t-tests, when looking at the influence of several interrelated factors on the outcome of an experiment (17). Two way analysis of variance, could be used to examine the effects and inter-relationship of pre-treatment condition score and concentration of glycerol on the success of treatment. The use

of this parametric technique requires that the data is measured on an interval scale. An adaptation of condition score is needed to produce data in this form.

Conclusion

It is common practice in conservation to record pre-treatment condition. Our research has demonstrated the potential of quantifying condition assessments in order to produce useful data about the treatment process. It is also necessary to produce a standardised method of defining the condition of objects for future studies and comparisons. This can be achieved merely by formalising the definitions of condition information, as recorded in the conservation documentation. By using a criterion anchored rating scale, a condition score can record specific and reliable information about object condition. This is a useful method of quantifying condition without the need for complex analytical assessments. The application of clinical methodologies to conservation questions allows us to refine and adapt treatments according to the demands of real treatment situations. In this study we have used the calculation of percentage treatment score from condition scores to enable us to assess the efficiency of the glycerol treatments and examine methods of determining their application.

In this study the use of condition score was found to be a successful method of allocating a concentration of glycerol solution to the individual treatment requirements of each artefact. The study also confirmed that material in poor condition should be treated with a higher percentage of glycerol, than material in good condition. Further work, based on a more rigorous experimental design, will attempt to verify this; identify the optimum range of glycerol concentrations; and examine the relationship of artefact condition to treatment requirements in more detail.

Acknowledgements

The authors would like to thank the conservation staff of the Museum of London and the Museum of London Archaeology Service for their help and comments. All our colleagues and the students from the Institute of Archaeology who helped us in collecting the data during the treatment process, and particularly Liz Pye, Clive Orton and John Mahonney Philips for their help and advice.

References

1. Ganiaris, H., Keene, S., and Starling, K., A comparison of some treatments for excavated leather. The Conservator 16 (1982) 12-23.

2. Chatfield, C. Statistics for Technology. 2nd Edition. Chapmans and Hall, London (1978).

3. Mills Reid, N., and Macleod, I. The use of aqueous glycerol solutions in the conservation of waterlogged archaeological organic materials. In Ambrose, W., and Mummery, J., eds. Archaeometry; Further Australasian studies (1987) 292-302.

4. Starling, K., The freeze drying of leather pre-treated with glycerol. ICOM 7th triennial meeting. Copenhagen (1984) 19-21.

5. Guidelines for the care of waterlogged archaeological leather. English Heritage (1995)

6. Crummy N. and Larkin D. Guidelines for the preparation of archaeological archives to be deposited with the Museum of London. Museum of London, forthcoming.

7. Ramachandran, G. N. , and Reddi , A.H. The biochemistry of collagen. Plenum Press New York (1976)

8. Mills Reid, N. Polyethylene Glycol Vs Glycerol. The importance of Molecular Structure IICM Bulletin, Issue 12 (1986) pp 51-61.

9. de Loecker, W., de Wever, F., Juliet, J., and Stas, M.L., Metabolic changes in rat skin during preservation and storage in glycerol buffer at -196 °C. Cryobiology 13 (1976) 24-30.

10. Hart, G.J., Russell, A.E., and Cooper, D.R., The effects of certain glycol, substituted glycols and related organic solvents on the thermal stability of soluble collagen. Biochemistry Journal 125 (1971) 599-604.

11. Larsen, R., Ed. STEP Leather Project; Evaluation of the correlation between natural and artificial ageing of vegetable tanned leathers and the determination of the parameters for the standardisation of an artificial ageing method. Protection and conservation of European Cultural Heritage Research report no.1 (1994)

12. Landy and Farr. The measurement of work performance, Academic Press, London (1983)

13. Keene, S., Real time survival rates for treatments of archaeological iron. In Proceedings of 1991 Conference on historic metals, Marina del Ray, CAL/ GCI. pp. 46-263, (1991)

14. Alderman M, An introduction to epidemiology. London Macmillan 1976

15. Sully, D. and Suenson-Taylor, K. A condition survey of glycerol treated freeze dried leather in long term storage. In Archaeological conservation and its consequences, Pre-prints of Triennial congress of the IIC, Copenhagen. pp. 177-181 (1996)

16. Siegal, S. and Castellan, N.J. Nonparametric statistics for the Behavioural Sciences 2nd Ed. McGraw-Hill, New York. (1989).

17. Clark , G. M. and Cooke, D. A basic course in Statistics. 3rd Ed. Edward Arnold, London (1992)