INTERNATIONAL COLLABORATIVE STUDY: STRUCTURING AND SAMPLE PREPARATION

G T COOK[1], D D HARKNESS[2], B F MILLER[2], E M SCOTT[3] M S BAXTER[1] and T C AITCHISON[3]

ABSTRACT. The success of any intercomparison exercise depends largely on participation and cooperation of a sufficient number of laboratories and the selection of a suitable suite of samples. Unless the latter is satisfactorily devised, the former cannot be guaranteed. The hierarchical nature of this study has necessarily resulted in a far more comprehensive set of sample types than has previously been employed. The exercise was structured to satisfy the following criteria: 1) to enable the participating laboratories to assess the experimental precision and accuracy of the component stages of the dating process; 2) samples should be typical of those routinely dated by the laboratories. This takes on a particular significance in Stage 1 where they should resemble as closely as possible the counting medium; 3) an objective statistical analysis of the results at each component stage of the study.

INTRODUCTION

Despite the almost universal use of NBS Oxalic Acid secondary standards (Oxalic I = SRM 4990, Oxalic II = SRM 4990-C) or a tertiary standard based on them, there have been previous suggestions that systematic biases can exist between ^{14}C laboratories (ISG 1982, 1983). Results have also been variable within laboratories. The partial failure of such a standard undoubtedly reflects the complexity of the dating process rather than any inherent inadequacy of the material itself. Nevertheless, this failure leads to a loss of resolution in dating, which consequently has led to reduced user confidence in the ^{14}C dating method. Inevitably, attention has focused on the comparability of conventional age measurements issued from different laboratories. Such suspicion then tends to grow to perhaps unrealistic levels when age measurements do not align with user preconceptions.

The previous practice of *ad hoc* exchange of samples among laboratories to satisfy internal and external confidence is now rightly perceived inadequate. In recent years, this has led to the initiation of a number of intercalibration exercises (Otlet *et al* 1980; ISG 1982). Although a previous study (ISG 1982) identified the existence of systematic bias and unexplained variability, its nature did not enable the identification of the component stage(s) in the overall technique which was responsible. In light of the discussion and consensus of opinion at the 12th International Radiocarbon Conference (Trondheim) in favor of a further and more comprehensive study, Scott *et al* (1986) devised a three-tier hierarchical study. First, if results again indicated a significant degree of variability between laboratories beyond what would be expected on the basis of quoted errors, the structuring might enable some progress towards identifying the sources of such variations and an assessment of their magnitude. Second, the use of several pairs of duplicate samples would enable estimation of both internal precision and the validity of the magnitude of routinely quoted error terms. Third, such a study would afford an opportunity for cross-calibration and verification of diverse laboratory procedures, particularly in light of the growth in application of accelerator mass spectrometry (AMS). Fourth, once the quantitative facts became known, the ^{14}C community might wish to act in concert to improve laboratory techniques and minimize sources of analytical uncertainty.

[1]Scottish Universities Research & Reactor Centre, East Kilbride, Glasgow G75 0QU, Scotland
[2]NERC Radiocarbon Laboratory, East Kilbride, Glasgow G75 0QU, Scotland
[3]Statistics Department, University of Glasgow, Glasgow G12 8QW, Scotland

STRUCTURING THE STUDY

The study was hierarchically structured in three-tiers to assess the existence and magnitude of any bias and additional variability in the following component stages in the overall dating procedure.

Stage 1 - Sample counting
Stage 2 - Sample synthesis
Stage 3 - Sample pretreatment

The success of any intercomparison exercise depends largely on two factors: 1) the participation and cooperation of a sufficient number of laboratories, and 2) the selection of a suitable sample suite.

Further, it is obvious that unless the latter is satisfactorily devised, the former cannot be guaranteed. Consequently, much deliberation centered on the choice of samples. We felt that desirable samples should represent, as far as possible, a reasonable cross-section of the types and ages of materials routinely analyzed; they should also be appropriate to the processes under consideration in each stage. The quality and isotopic homogeneity of each batch of material was naturally of paramount importance.

Stage 1

The specific objective here was to assess as closely as practicable the degree of bias and of internal and external precision derived from the counting procedure alone. For the liquid scintillation (LS) laboratories, the obvious choice of material was benzene. Besides being the ubiquitous counting medium for LS laboratories, benzene has the following advantages (Otlet *et al* 1980): 1) it is a basic hydrocarbon produced via petrochemical processing to high purity and of infinite ^{14}C age, and 2) ^{14}C-labeled solutions of differing activities may be prepared homogeneously by simple dilution and mixing.

For the gas proportional counting (GPC) and accelerator mass spectrometry (AMS) laboratories, the selection of material was much less obvious. A variety of counting gases are employed by the GPC laboratories, *viz*, CO_2, C_2H_2 and CH_4, whereas the requirement of extremely high purity, especially for CO_2 laboratories, would probably have necessitated further purification upon receipt. Also, problems of transport would be prohibitive and target production to the almost unique specification of each AMS laboratory would have placed an insurmountable constraint on the sample preparation program. We eventually decided that calcium carbonate represented the best compromise for both techniques - barium carbonate had to be discounted because of shipping safety rules governing barium compounds. Although a range of benzene activities could have been produced to a high degree of precision, a similar suite of $CaCO_3$ samples would not have been possible. At any rate, we considered it more important that the samples should enable estimates of: 1) the existence of systematic biases among laboratories; 2) the validity of routinely quoted errors; 3) internal precision by comparing the difference between duplicates with their quoted errors. Consequently, Stage 1 was based on a total of four bulk samples. Each LS laboratory received four samples of benzene, *viz*, duplicates of ca 100% modern and duplicates of ca 50% modern (Fig 1). We sent a similar suite of $CaCO_3$ samples to the AMS and GPC laboratories. No indication was given as to the existence of duplicates. Harkness *et al* (1989) present a full account of the preparation of these materials.

Liquid Scintillation
Laboratories

Gas Proportional Counting/
Accelerator Mass Spectrometry
Laboratories

Benzene

$CaCO_3$

duplicates
ca 100%

duplicates
ca 50% modern

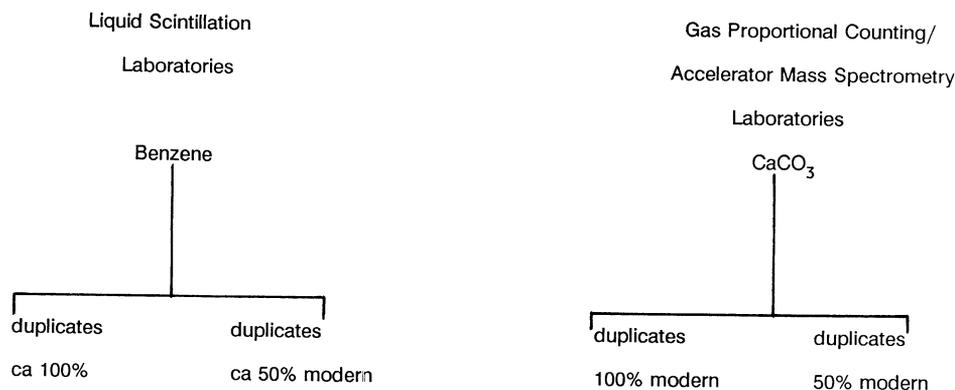duplicates
100% modern

duplicates
50% modern

Fig 1. Graphic representation of Stage 1

Stage 2

We envisaged that samples for Stage 2 should encompass both syntheses to the counting medium and the counting process itself. Thus, any variability additional to that observed in Stage 1 could be attributed to the synthesis component. Three sample types were available and again, each was supplied in duplicate following rigorous pretreatment:

1.   A bulk sample of marine algae (*Lithothamnion*) collected from a naturally occurring deposit in Argyll, Scotland. The fine granular structure of this material lent itself to thorough mixing at various stages in the pretreatment schedule to ensure homogeneity.

2.   A bulk humic acid extract was obtained from a depth interval, ca 20cm thick, of well-humified peat from central Scotland. We preferred the humic acid fraction because it is in solution at an intermediate stage in the extraction procedure. At this point it was possible to bulk and thoroughly mix the solutions prior to precipitation, thereby guaranteeing homogeneity. Because one of the samples in Stage 3 was the parent peat material, additional assessment is possible.

3.   The University of Belfast generously provided a large sample of dendrochronologically dated bog oak comprising 20 years' growth (241-260 BC). This was finely chopped and rigorously pretreated to yield a final fibrous cellulose product. Harkness *et al* (1989) give a full account of the pretreatment and homogenization of the Stage 2 samples.

One criticism of the Stage 2 materials was voiced at the 13th International Radiocarbon Conference (Dubrovnik) - that selected pieces of the marine algae showed larger-than-expected discrepancies in age when dated by AMS, suggesting that the bulk sample did not represent a single event. However, the overall results of the analyses of duplicates do not suggest any poorer internal precision, nor do they reveal any evidence of greater external bias. We must then begin to consider whether this criticism reflects a problem of sample definition. The algal material was identified and collected as a naturally occurring sample typical of that which might be submitted to a ^{14}C laboratory for routine analysis. The pertinent question then is how AMS practitioners choose to define and address representative sampling; whether this be pre- or post-receipt of materials selected for ^{14}C measurement. In this particular instance, our attention is drawn to a prime example where deliberate small-scale preselection from a typical natural deposit can enhance

the possibility of a significant deviation from its most representative age. Of course, the importance of representative sampling is by no means a novel concept for ^{14}C dating. However, there is clearly a need for particular awareness in AMS laboratories and indeed others which assay micro-samples.

Stage 3

We selected the final suite of samples to examine the entire dating process. We agreed that the sample types should reflect the pretreated materials used in Stage 2, while at the same time, a relatively high proportion should be absolutely dated materials. Thus, Stage 3 comprises 8 samples in total: duplicate samples of peat from the original bulk sample used to prepare the Stage 2 humic acid extract, duplicate samples of bivalve mollusks (*Anadara antiquata*) and finally 4 wood samples made up of duplicate samples of dendrochronologically dated wood (dendro-age 221-240 BC) and 2 single samples of dendro-dated wood, AD 1521-1550 and AD 1841-1870. Using known-age materials allows comparison of the performance of individual laboratories with respect to an absolute age value rather than to a measure of a central value. Also, we can assess any bias that the study group shows as a whole.

REFERENCES

Harkness, DD, Cook, GT, Miller, BF, Scott, EM and Baxter, MS 1989 Design and preparation of samples for the international collaborative study. *In* Long, A and Kra, RS, eds, Internatl ^{14}C conf, 13th, Proc. *Radiocarbon* 31(3): 407-413.

International Study Group 1982 An inter-laboratory comparison of radiocarbon measurements in tree-rings. *Nature* 298: 619-623.

_____ 1983 An international tree-ring replicate study. *In* Waterbolk, HT and Mook, WG, eds, Proc. Strasbourg, *Pact* 8: 123-133.

Otlet, RL, Walker, AJ, Hewson, AD and Burleigh, R 1980 ^{14}C interlaboratory comparison in the UK: Experiment design, preparation and preliminary results. *In* Stuiver, M and Kra, RS, eds, Internatl ^{14}C conf, 10th, Proc. *Radiocarbon* 22(3): 936-947.

Scott, EM, Baxter, MS, Aitchison, TC, Harkness, DD and Cook, GT 1986 Announcement of a new collaborative study for intercalibration of ^{14}C dating laboratories. *Radiocarbon* 28(1): 167-169.