AN INTERIM PROGRESS REPORT ON STAGES 1 AND 2 OF THE INTERNATIONAL COLLABORATIVE PROGRAM

E M SCOTT**, T C AITCHISON*, D D HARKNESS**, M S BAXTER†

and

G T COOK†

ABSTRACT. We report in this paper on a preliminary analysis of Stages 1 and 2 of the International Collaborative program. We have chosen to concentrate on the internal and external consistencies of the participating laboratories. The two stages so far completed have dealt only with the processes of sample synthesis and counting, and results indicate that the major component of variability lies in the counting process. Outlying laboratories are observed at each stage. A third stage is in progress which will allow an assessment of any further variability due to sample pretreatment. With the inclusion of duplicate samples in each stage, we are able to report that laboratories are remarkably consistent internally, ie, the differences between duplicates generally agree with the laboratory's claimed precision.

INTRODUCTION

The ^{14}C dating community has acknowledged the importance of inter-laboratory checks through its willingness to participate in a number of collaborative studies (Otlet et al, 1980; ISG, 1982, 1983). Currently, a third and considerably more ambitious project (Scott et al, 1986) has been initiated and is being supported by over 50 ^{14}C laboratories. We report here on the progress of this third study.

Overall Aims

The study has been undertaken with the aims of 1) gaining insight into the contribution of the various dating processes to the overall dating error, 2) providing experimental verification of the diverse laboratory techniques used in dating, and 3) quantifying uncertainties on routine results obtained by the modern generation of ^{14}C laboratories.

Organization and Design of the Study

Full details of the study organization are given in Scott et al (1986). The study has three sequential stages, each introducing a further laboratory process into the study. Stage 1, primarily involving the counting process, was completed and reported on in September 1987 (Scott et al, 1987). Stage 2, which introduced sample synthesis, was completed in April 1988 and we report our preliminary findings here. Stage 3, which introduces sample pretreatment, is currently underway and will not be completed before January 1989.

* Dept of Statistics, University of Glasgow, Glasgow G12 8QW, Scotland
** NERC ^{14}C Laboratory, NEL, East Kilbride, Glasgow, G75 0QU, Scotland
† Scottish Universities Research and Reactor Centre, East Kilbride, Glasgow, G75 0QU, Scotland

414

STUDY SAMPLES

The hierarchical sample structure is an important element of the study design. Table 1 illustrates the sample types offered in stages throughout the program. Duplicate samples were included at each stage as well as four known-age samples in Stages 2 and 3. In total, each laboratory participating in all three stages will complete a minimum of 16 ^{14}C dates. A full description of the sample preparations is given by Harkness *et al* (1989).

Eighty laboratories were invited to participate, of which 58 full participants were registered. A total of 52 laboratories returned results for Stage 1 and then received Stage 2 samples. Of these, 37 have, thus far, returned results for Stage 2 and received Stage 3 samples. Table 2 shows the number of laboratories involved and the composition of the study group at various stages, in terms of counting technique and medium.

TABLE 1

Sample types

Stage 1 – Calcium carbonate or benzene
Stage 2 – Humic acid, cellulose, algal lithothamnion
Stage 3 – Shell, peat and wood

TABLE 2

Breakdown of laboratory participation

Stage	Gas counting			LSC	AMS
	CO_2	C_2H_2	CH_4		
1	19	1	4	20	8
2	13	1	4	14	5

STAGE 1 RESULTS

We presented an analysis of the results from Stage 1 in Groningen (Scott *et al*, 1987) and, therefore, we only summarize our findings here.

The main aim of Stage 1 was to quantify variability in results due to the counting process and, to this end, samples in Stage 1 were either benzene or calcium carbonate. Within each sample type, two distinct '^{14}C ages' were provided and each laboratory received two duplicate samples of each age. The laboratory did not know the identity of each sample. Table 3 shows the summary statistics of the results for the four distinct samples.

TABLE 3

Summary statistics for Stage 1

Sample	Carbonate 1	Carbonate 2	Benzene 1	Benzene 2
n	32	32	20	20
Average	−66	3600	930	6900
Range	500	850	1300	1700
Quartile Range	−110–0	3570–3670	705–1060	6760–7100

From the four results from each laboratory, the average of, and the disparity between, each duplicate pair was calculated. Disparity is defined as the difference between duplicate pairs (ignoring sign), divided by the square root of the sum of the errors squared. The disparity is a measure of how much the laboratory quoted errors relate to the observed difference between the duplicate results. Values of the disparity exceeding two would indicate that a laboratory was unable to detect the presence of duplicates within the limits of their claimed precision.
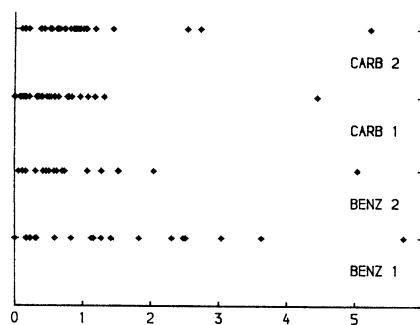

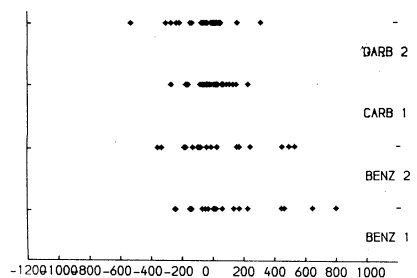Fig 1. Disparities for all sample types in Stage 1

Fig 2. Average of duplicates for each sample in Stage 1 relative to the overall sample average for each sample type

Figure 1 shows the disparities for the four samples of Stage 1. Outlying laboratories (a disparity exceeding two) are shown clearly in this figure and are more pronounced for the first benzene sample, which was ca 1000 yr BP.

Seven out of 20 liquid scintillation laboratories had disparities, on 1 or other of the samples, which exceeded 2. This effect is particularly pronounced on the younger sample. Only 4 out of 32 gas counting and accelerator laboratories had disparities exceeding 2 and there was no evidence that accelerator laboratories are separable, in this aspect of quality, from the gas counting laboratories.

Figure 2 shows the variability in the average results for each Stage 1 sample. Here, the overall average for the sample has been subtracted from each individual laboratory average. Again, there is some evidence that the data from the liquid scintillation laboratories are more widely scattered than those from the gas counting and accelerator laboratories (the range of results for the liquid scintillation laboratories is almost twice that of the gas counting and accelerator results). Such a difference, however, results from one or more extreme outliers in the liquid scintillation data set.

Considering the relationship between the averages of the duplicate pairs gave a highly significant correlation coefficient of $r = 0.78$, for those laboratories that assayed benzene. Such a high degree of linear dependence is evidence for systematic biases. For those laboratories that analyzed the carbonate samples, the correlation was much lower at $r = 0.37$, indicating a less clear linear dependence between the results. Again, the influence of outlier laboratories was indicated.

Currently, to help assess the experimental causes of the observed distribution of results, additional information is being sought from participating laboratories, in particular, concerning reference standards.

STAGE 2 RESULTS

Stage 2 introduced the further process of sample synthesis into the study. Bulked homogenized samples of cellulose, humic acid and algal lithothamnion were prepared and distributed. Laboratories selected 2 of the 3 sample materials and received 2 duplicate samples of each.

Table 4 summarizes results received thus far from 37 laboratories. It is clear that the ranges of results for both the cellulose and algal samples are considerably larger than that for the humic acid sample but that the mid

TABLE 4

Summary statistics for Stage 2

Sample	Humic acid	Cellulose	Algal lithothamnion
n	17	36	24
Average	3390	2250	2120
Range	550	1900	1800
Quartile Range	3920–3450	2160–2370	2040–2200

50% ranges (the difference between the quartiles) for each sample type are roughly equivalent at 200 yr. These statistics again indicate one or more extreme laboratories in the group. Evidence of lack of homogeneity was found in the natural algal sample by the accelerator laboratories but was not detected by the conventional laboratories.

Also, we analyzed various sample type combinations. The three possible sample type combinations are humic acid/algal, humic acid/cellulose and cellulose/algal. Four, 16 and 23 laboratories respectively opted for these possible combinations. An assessment of the dependence between the

average results for each sample type combination again showed high degrees of linear dependence (typically correlations of 0.8 or higher). This again provides evidence of systematic biases.

Analysis of the average of and disparity for each duplicate pair showed that most laboratories have disparities <2, indicating that, internally, laboratories are remarkably consistent, *ie,* they are able to detect duplicates within the limits of their claimed errors. In general, small differences between gas counting and accelerator and liquid scintillation laboratories are not apparent, although there is again some indication of more variability in the results reported by liquid scintillation laboratories.

Known-age material was included in this stage (cellulose samples). The absolute date of the wood sample provided by the Belfast dendro-
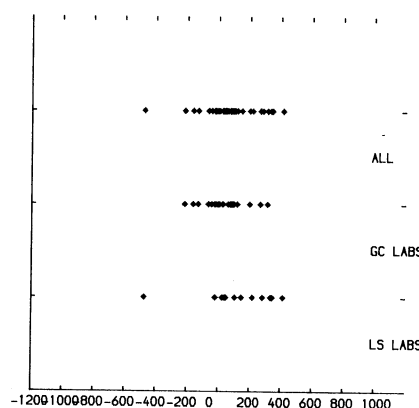

Fig 3. Average of duplicates for cellulose sample in Stage 2 relative to known age

chronological laboratory and from which the cellulose was extracted was 241–260 BC. Using the calibration data of Pearson *et al* (1986), these calendar dates would correspond to ^{14}C ages of 2220–2160 BP. Figure 3 shows the relative offsets to a nominal ^{14}C age of 2180 BP for liquid scintillation, gas counting and accelerator laboratories. There is some indication of a small systematic deviation from the perceived true age. Additional known-age samples included in Stage 3 will allow further comparison between the ^{14}C dates and absolute base lines and will, hopefully, answer questions on laboratory biases.

STAGES 1 AND 2

It is of considerable interest to compare the results from Stages 1 and 2 to evaluate the impact of sample synthesis on the observed variability. Figures 4 and 5 present the laboratory offsets and disparities for liquid scintillation laboratories for the five samples of Stages 1 and 2. Again, outlying
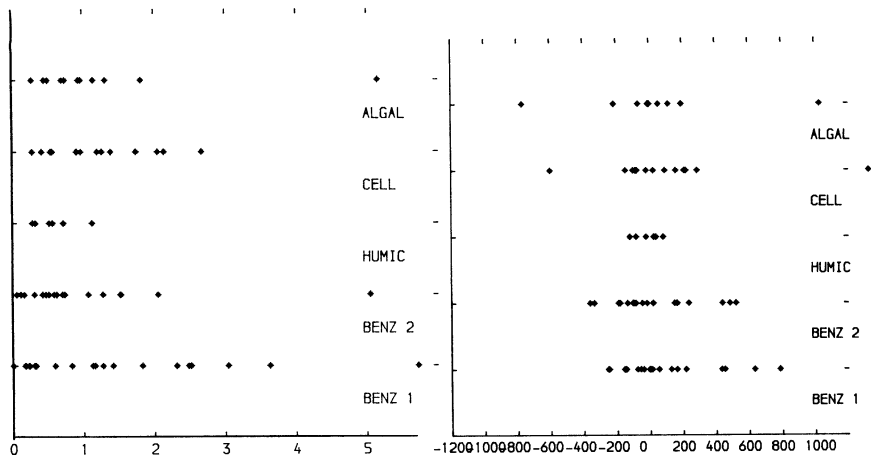
Fig 4. Disparities for Stages 1 and 2 for liquid sci-
ntillation laboratories

Fig 5. Averages of duplicates for Stages 1 and 2 for
liquid scintillation laboratories relative to the
overall average for each sample type

laboratories are highlighted however Stage 2 shows some reduction in data
scatter relative to Stage 1.

Figures 6 and 7 present the same quantities for the gas counting and
accelerator laboratories. Here, there is little apparent difference between
Stages 1 and 2, which is perhaps, to be expected since, in both stages, the
samples had to be converted to the chosen counting materials.

Overall, there is some support for the contention that the major com-
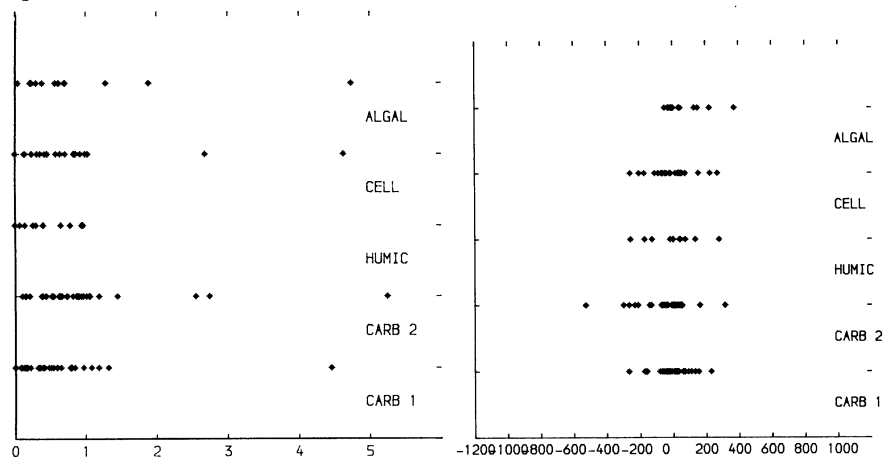ponent of variability in the results derives from the counting process.

Fig 6. Disparities for Stages 1 and 2 for gas counting
and accelerator laboratories

Fig 7. Averages of duplicates for Stages 1 and 2 for
gas counting and accelerator laboratories relative to
the overall average for each sample type

CONCLUSIONS

The preliminary findings of this study deal mainly with the internal and the external consistencies shown by the participating laboratories. Several general points are also worth making.

First, it is clear that, at each stage and for each laboratory type, there are outlying laboratories (defined either in terms of disparity or laboratory offset). The analysis presented here has not removed the data for any of these laboratories, although clearly results could be improved by their exclusion. These laboratories are being, and have been, asked to provide further experimental information and to re-examine their procedures.

Second, results show clearly that laboratories are remarkably consistent internally, *ie*, the differences between the duplicates generally agree with the laboratory's claimed precision.

Third, the study indicates systematic biases amongst the laboratories and that there seem to be problems in calibrating to a common standard. Indeed, we see no improvement from the previous study (ISG, 1982, 1983).

Fourth, our preliminary findings from Stages 1 and 2 suggest that the major component of variability lies in the counting process. Sample synthesis appears not to have introduced further significant sources of variability; however, for gas counting and accelerator laboratories, Stage 2 did not introduce any additional processes.

Thus, we await the results from Stage 3 to complete the analysis and to indicate the effect of sample pretreatment on the scatter of results.

FUTURE DEVELOPMENTS

There is a clear need for regular intercomparison exercises to provide opportunities for laboratories to cross-check their results. Future studies need not be as large scale as this study and should introduce further variety in sample types and age range of samples on offer. It is, however, important that realistic sample materials be used to ensure the applicability of the findings to the routine operation of the laboratory.

We hope to convene a workshop in Glasgow in 1989 to discuss further developments from this and other work (see Scott, 1989).

ACKNOWLEDGMENTS

REFERENCES

Harkness, DD, Cook, GT, Miller, BF, Scott, EM and Baxter, MS, 1989, Design and preparation of samples for the International Collaborative Study: Radiocarbon, this issue.

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

————— 1983, An international tree-ring replicate study, *in* Internatl symposium, ^{14}C and archaeology, 1st, Proc: Strasbourg, PACT 8, p 123–133.

Otlet, RL, Walker, AJ, Hewson, AD and Burleigh, R, 1980, ^{14}C interlaboratory comparisons in the UK: Experiment design preparation and preliminary results, *in* Stuiver, M and Kra, RS, eds, Internatl ^{14}C conf, 10th, Proc: Radiocarbon, v 22 no. 3, p 936–946.

Pearson, GW, Pilcher, JR, Baillie, MGL, Corbett, DM and Qua, F, 1986, High precision ^{14}C measurement of Irish oaks to show the natural ^{14}C variations from AD 1840 to 5210 BC, *in* Stuiver, M and Kra, RS, eds, Internatl ^{14}C conf, 12th, Proc: Radiocarbon, v 28 no. 2B, p 911–934.

Scott, EM, 1989, Report on the Workshop on the International Collaborative Study: Radiocarbon, this issue.

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, v 28, no. 1, p 167–169.

Scott, EM, Baxter, MS, Harkness, DD, Aitchison, TC and Cook, GT, in press, Recent progress in the International Calibration of Radiocarbon Laboratories, *in* Internatl symposium, Archaeology and ^{14}C, 2nd, Proc: PACT.