

**¹⁴C INTERLABORATORY COMPARISON IN THE UK:
EXPERIMENT DESIGN, PREPARATION AND
PRELIMINARY RESULTS**

R L OTLET*, A J WALKER**, A D HEWSON***, and
R BURLEIGH***

ABSTRACT. An interlaboratory comparison experiment for ¹⁴C measurements has been organized jointly by the Harwell and British Museum laboratories to include the working radiocarbon laboratories of the United Kingdom. The experiment has been run along the lines of that organized by the IAEA for tritium and has explored the problems of sample preparation, verification of equivalent levels, and presentation of results. Samples of benzene representing 5 age equivalent levels between twice modern and 20,000 years old have been prepared and distributed for measurement either by direct counting (liquid scintillation) or full process tests. Preliminary results received show excellent agreement both between laboratories and in comparison with the known relative activities of the prepared solutions. The possibility of extending the experiment to cover different sample types and a wider distribution of testing laboratories is briefly discussed.

INTRODUCTION

For some years it has been felt that an interlaboratory comparison project similar to that organized by the IAEA for low-level tritium measurements (Florkowski, Payne, and Sanzay, 1970) would be a valuable demonstration of the consistency of carbon-14 measurements from laboratories world-wide. At the modern level, the calibration is generally considered to be adequately controlled by the use of the common standard, oxalic acid, produced and issued by the National Bureau of Standards. Other possibilities for modern standards have also been discussed (Polach, 1976) and the analysis of the international cross-calibration data was discussed in the previous paper. However, in extrapolating calibration throughout the usable time-scale afforded by carbon-14 and over a wide range of sample sizes, the question has remained open as to whether interlaboratory alignment is precisely maintained. Undoubtedly, many *ad hoc* exchanges of samples between laboratories have taken place. A number of detailed comparisons of inter-laboratory results have also been made, eg, by Polach (1972) who evaluated replicate oxalic acid measurements, and by Clark (1975) in his analysis of bristlecone pine measurements.

In some disciplines, full interlaboratory comparisons are regularly undertaken. The logistics of running such an experiment for ¹⁴C are rather daunting. The experiment reported here, a collaborative venture undertaken jointly by the Harwell and British Museum laboratories, aims to provide an insight into the possible outcome of undertaking such a full interlaboratory comparison from the designing and running of a specific experiment on a less ambitious scale.

Nine laboratories were invited to take part and, at the time of writing, results, although not complete, have been received from 8 of them. All laboratories were asked to treat these samples in their normal routine

* UKAEA Harwell
** UKAEA Harwell and University of Bradford
*** British Museum

manner and an assurance was given at the outset that strict anonymity would be preserved in the final presentation of results.

Experiment design

Five levels of ¹⁴C activity were selected to represent the time period under examination. These were equivalent to radiocarbon ages of 20,000, 10,000, 5000 and 2000 years and present day (~1963) peak level (200% modern). A further sample expected to be of infinite age and which had been used to dilute the other samples was also included.

In the original suggestions for the experiment it was intended that three types of samples should be prepared and distributed. These were benzene, specifically to test the counting precision of liquid scintillation laboratories, oxalic acid for chemical processing plus counting by all laboratories and, as the final stage, natural known-age samples to test the complete process. Up to now, only benzene has been considered but, as will be shown, this has been used beyond just the testing of the counting of liquid scintillation systems. By combusting samples to CO₂ and following normal measurement procedure processes from thereon, full process tests have been possible in some cases and these are also included in the final results.

Benzene has a number of specific advantages as the choice of sample for an intercomparison exercise.

- 1) It is a basic hydrocarbon produced petrochemically (hence, of infinite ¹⁴C age) and available commercially at a high purity level.
- 2) ¹⁴C labelled solutions can be prepared homogeneously by simple mixing.
- 3) Dilution to different levels can be achieved accurately by weight (and prepared levels can be checked against the weighing by direct liquid scintillation counting).
- 4) The initial ¹⁴C active benzene used for the labelling is conveniently prepared to high purity using the established benzene synthesis process.

This last point is important because in all labelled activity experiments the radioactive component, should be of at least the same purity as the diluting medium, however small a proportion it forms of the final mixture.

The labelling benzene used in this case, ~200 mls of ~100 × modern solution, was a dilution of benzene synthesized at a very high level in the course of a previous laboratory experiment.

Rigorous tests were made to check the purity and absence of ¹⁴C in the 'dead' benzene used for the subsequent dilutions. First, from one of a common batch of 22+L of Analar grade benzene (9 × 2.5L bottles) samples of varying weights were dispensed for counting following the normal procedures of this laboratory (Otlet and Warchal, 1978). Counting efficiencies indicated by the external source ratio (Packard AESR) were compared with mean values determined for samples of similar weight groupings synthesized in the course of normal dating work over

TABLE I
Qualitative comparisons of quenching from ANALAR and laboratory synthesized benzene

Benzene (g)	External source ratio (and counting efficiency (%))	
	ANALAR	Synthesized*
3	0.8721 (75.8)	0.8751 (75.8)
4	0.8612 (75.8)	0.8608 (75.8)
5	0.8411 (75.7)	0.8419 (75.7)
6	0.8290 (75.6)	0.8155 (75.5)
7	0.8078 (75.5)	0.8012 (75.3)

* From analysis of 150 samples counted over a period of 12-18 months.

the previous 12 to 18 months. As seen in table 1, no significant differences were observed. Further verification that the purity was well up to the standard of benzene synthesized for normal liquid scintillation procedures (Fraser and others, 1974) was provided by tests on an Elmar gas/liquid chromatograph. Traces of toluene, acetone (or possibly n-hexane) and ethanol were observed, in ratios 4:2:1, respectively, but at levels well below 1000 ppm toluene.

Secondly, the ^{14}C level of samples from each of the nine bottles of 'dead' benzene was compared with established background standards generally used by the organizing laboratories. No significant differences were observed between the samples and backgrounds, thus confirming the suitability of the selected AR benzene for the dilutions. The nine bottles were then tipped into a 25L carboy, thoroughly mixed to guarantee homogeneity, and redispensed ready for use into the original 2.5L bottles.

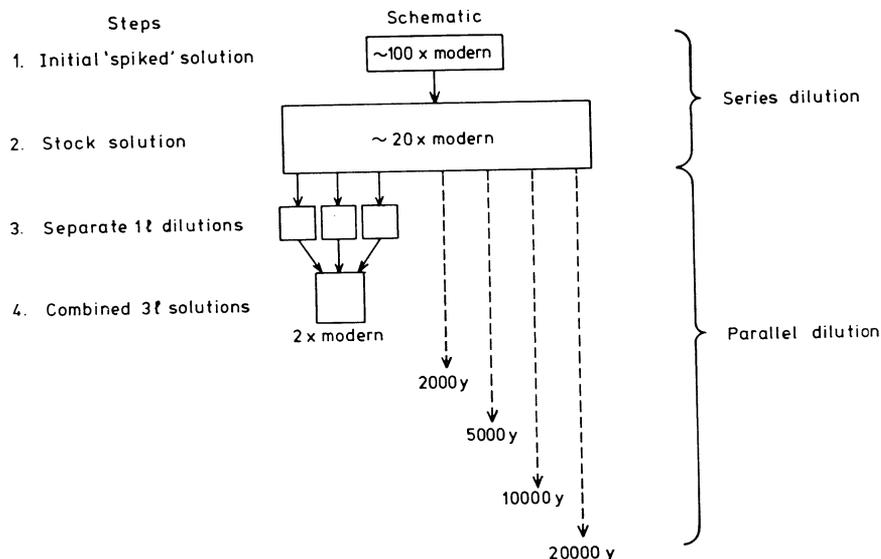


Fig 1. Dilution procedures.

Dilution procedure

Two methods of dilution were considered, a series and a parallel system. In a series system aliquots from one level are used with dilution to produce the next level down, and so on. This has the advantage of maintaining only small dilution ratios throughout but allows errors in any step to accumulate to a maximum in the final level. In the parallel method each level is diluted independently. Errors tend to be larger as the dilution ratios increase but this can be largely offset by carrying out the dilutions by weighing with a suitably precise balance. The initial level is then chosen such that the dilution to the lowest level is still within the working range of the balance. A combination of these methods was adopted as shown in figure 1, using an initial series dilution from 100 times 'spike' to ~20 times modern stock solution and a parallel system from thereon.

With the balance available, Sartorius type 370 MP1, a precision of ± 0.01 g was possible but its working range limited the maximum volume of any operation to ~1L. Thus, to achieve a final quantity of 3L at each equivalent age level, three separate, identical dilutions were required. Before combining the 3 separate 1L dilutions, 3 samples were taken and counted from each flask. A summary on the counting of these 45 samples is given in table 2. (A previous test on 9 samples from the ~20 times modern stock solution (table 3) confirmed the precision of the counting techniques used and provided a mean value of the activity level from which to plan the subsequent dilutions).

Analysis of variance on the data of table 2 and on the individual counting residuals (not shown) suggested no differences at the 5 percent significance level either between samples or between flasks with one possible exception, group E. A second test of 5 samples from each of the E flasks, with longer counting on each sample than in the initial tests,

TABLE 2
Summary of counting on the initial separate dilutions prior to
combination into 3 liter lots at each level

Flask no.	Sample no.	Levels (values in % modern)				
		A	B	C	D	E
1	1	196.50	77.30	52.05	28.97	9.04
	2	197.74	77.25	53.11	29.33	8.53
	3	198.97	77.01	53.10	29.05	8.38
2	1	196.52	76.95	53.46	29.07	7.90
	2	199.02	76.85	53.25	29.46	7.95
	3	198.26	76.69	53.99	28.87	8.32
3	1	196.05	76.80	53.58	28.82	9.28
	2	195.76	78.93	53.77	28.79	8.42
	3	197.01	77.67	52.51	28.36	8.87
Grand means and precision (standard errors)		197.31 \pm 0.41	77.27 \pm 0.23	53.20 \pm 0.20	28.97 \pm 0.11	8.52 \pm 0.16

later showed that this group too was satisfactory. Accordingly, the three flasks of each group were combined forming approximately 3L total solution at each activity level.

Intercomparison procedure

Samples were first distributed to all those ^{14}C laboratories in the UK able to make direct liquid scintillation counting measurements. Each sample was labelled with a code letter denoting the activity level and a number referring to its order of make-up from the stock solution, but the numbers sent to any one laboratory were deliberately non-serial. Samples were weighed into the bottle (30ml, ~26g) and each weight written on the label so that the recipient could check that only insignificant evaporation had occurred in transit.

No information was given regarding the activity distributions except an approximate guide to the highest level (~ twice modern) in order to allay fears that contamination from an unknown high level could occur, and the order of the levels, *ie*, highest A in descending activity to E. It was also stated that sample F had been taken from the bulk solution used to dilute all the prepared samples. Confirmation of the original assumption that this solution was 'dead' benzene was clearly of value since this was crucial information in calculating the dilution ratios.

Secondly, for laboratories prepared to undertake full process tests (and this of course, included those operating gas counters), quantities of CO_2 were prepared at Harwell. Combustion of liquid benzene was carried out using a modified tube combustion line (fig 2) adapted from an original scheme provided by M Baxter (Glasgow University, pers commun) developed for the combustion of alcohol.

Aliquots of each activity level were combusted, and quantities of CO_2 varying from 5 STP liters to 10 STP liters were issued to each laboratory depending upon their particular process requirements. The distribution

TABLE 3
Counting reproducibility and establishment of the stock solution level

Sample no.	Benzene weight (g)*	AES ratio	Counting efficiency (%)	Nett** count rate (cpm)	Activity level (% mod)†
I37	4.2612	0.9157	74.64	775.58	1992.24
I38	4.3799	0.9127	74.67	798.01	1993.50
I40	4.5360	0.9004	74.76	826.87	1992.11
I41	4.3884	0.9075	74.72	799.53	1992.10
I42	4.4254	0.8949	74.79	805.92	1989.37
I43	4.4392	0.9031	74.75	809.12	1992.12
I44	4.4130	0.9052	74.73	805.80	1996.26
I45	4.4096	0.9231	74.55	802.64	1994.77
I46	4.3866	0.9097	74.70	799.91	1994.40

Overall result 1993.0 ± 0.7 (standard error) percent modern.

* Mean sample weight 4.4044 ± 0.07 (standard deviation)

** After background of 10.54cpm subtracted. Each sample initially counted to 400,000 total counts.

† Modern defined in terms of $0.95 \times$ the mean activity of 5 NBS oxalic acid samples counted simultaneously, 12.24 dpm/g C_6H_6 .

of gaseous CO_2 by the issuing laboratory is not recommended for future experiments since this does not test fully the measurement laboratory and it is also difficult to achieve without mishaps.

Laboratories were asked to report their results either as d^{14}C or as conventional radiocarbon ages bp. In accordance with the definitions of d^{14}C given by Stuiver and Polach (1977), reproduced in Appendix 1, correction for fractionation is applied to the oxalic acid standard but not to the sample count. Similarly, it was asked that no allowance for sample fractionation should be made in the age bp calculation.

It was further requested that as much basic data as possible should be provided regarding the counting system and sample count details.

- 1) The specific associated background and modern calibration readings.
- 2) Counting efficiency.
- 3) Total counts (or nett cpm and time of count).
- 4) Order of counting of the samples if these were not given a single continuous count.
- 5) Type of liquid scintillation spectrometer and channel settings (for details of gas counters).
- 6) Quantity of sample used.
- 7) Vial and 'cocktail' details.

Also, where more than one sample had been measured, individual results were requested rather than means.

Preliminary results and comments

Results are presented in table 4(A) (Dispensing and LS counting) and 4(B) (Full process measurements from CO_2). Both to preserve the strict anonymity promised to the participating laboratories and, in view

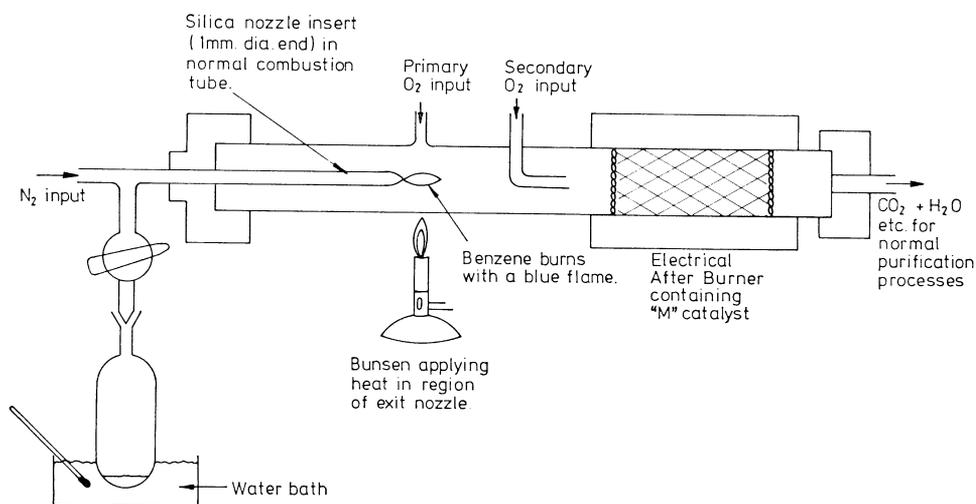


Fig 2. Benzene combustion apparatus.

of the preliminary nature of some of the estimates, it was decided not to publish the error terms of the measurements at this stage. Only the percent modern or equivalent age results are given.

The close agreement of all the results to one another, even by visual inspection, is immediately impressive. So, too, are the values of the one sigma standard deviations calculated directly from the results. All values are included, irrespective of their derivation (eg, 'same sample counted twice' etc) and no weighting factors have been applied. It is noteworthy that these values and particularly the standard deviations on the all-in pooled values given in table 5(i) are in good agreement with the kind of estimates quoted for actual radiocarbon determinations of comparable age.

In addition to the 'all-in' pooled means calculated without weighting table 5 includes, for comparison, means calculated using the weighting and test procedures of Ward and Wilson (1978). In their method weighting factors are calculated from the one sigma error terms (σ)

TABLE 4
Preliminary results

A. Dispensing and counting only

Laboratory codes (a)	Results for each coded level				
	A	B	C	D	E
1	200.0	1,990	5,100	10,150	20,410
2	198.0	2,130	4,950	9,950	19,740
3*) (a)	198.4	2,100	5,030	10,000	19,630
) (b)	197.0	2,190	5,080	10,160	19,810
4**) (a)	197.9	2,045	5,035	10,040	19,900
) (b)	198.3	2,105	5,160	10,100	19,790
5†) (a)	196.5	2,085	5,074	9,860	19,870
) (b)	200.6	2,090	5,000	9,900	20,090
6**) (a)	197.7	2,104	5,075	10,080	20,020
) (b)	197.0	2,081	5,089	10,046	20,009
Means and std devs	198.1 ± 1.3	2,092 ± 52	5,059 ± 58	10,029 ± 101	19,927 ± 220

B. Full process measurements from CO₂

Laboratory codes (b)	Results for each coded level				
	A	B	C	D	E
1	197.5	2,130	5,000	9,840	19,550
2†) (a)	198.7	2,070	5,030		20,320
) (b)	198.9	2,105	5,070		19,750
3*) (a)	198.4	2,190	5,070	10,050	19,870
) (b)	197.0	2,100	5,000	10,200	
4	198.1	2,050			20,180
Means and std devs	198.1 ± 0.7	2,108 ± 49	5,034 ± 35	10,030 ± 181	19,934 ± 314

* From two separate groups of samples counted on different LS spectrometers.

** From two separate groups of samples counted on the same LS spectrometer at different times.

† From the same group of samples counted on different LS spectrometers.

NB: Results for A are given as '% modern', others in years bp.

quoted by the originating laboratories and have values of $\frac{1}{\sigma^2}$. A pooled mean (A_p) is calculated from the individual results (A_i) such that

$$A_p = \frac{\sum_1^n \frac{A_i}{\sigma_i^2}}{\sum_1^n \frac{1}{\sigma_i^2}}$$

where n is the number of individual results.

Acceptance of the data is judged from the test statistic T, given by

$$T = \sum_1^n \frac{(A_i - A_p)^2}{\sigma_i^2}$$

(T has a chi-square distribution on n – 1 degrees of freedom).

Application of this procedure to the full results of the participating laboratories showed that the numerical value of T was in all cases close to the appropriate number of degrees of freedom and no data were therefore rejected. This might also be interpreted to mean that the errors quoted by the individual laboratories are good estimates of the true errors of their measurements.

Also in table 5 is a comparison of the dilution ratios (with respect to level A) derived from the pooled means, with those determined from the weighings of the actual dilution procedure. It is very satisfactory to observe that in all cases the agreement is closer than 1 percent.

TABLE 5
Comparison of pooled means with the experimental and derived dilution factors

Solution levels	A	B	C	D	E
(i) Pooled means and 1σ standard deviations (no weighting)	198.1 ± 1.1	2098 ± 50	5051 ± 52	10,029 ± 115	19,929 ± 244
(ii) Pooled means and 1σ standard errors (with weighting)	198.2 ± 0.23	2086 ± 8	5059 ± 12	10,054 ± 20	19,950 ± 43
(iii) Derived dilution factors relative to A (from data in (ii))	1	2.570	3.721	6.929	23.751
(iv) Experimental dilution factors relative to A	1	2.551	3.709	6.916	23.879
(v) Differences in (iii) wrt (iv) (%)	—	+0.7	+0.3	+0.2	−0.5

Conclusions and proposals for future tests

Although the results and their treatment are preliminary, certain valuable conclusions can already be drawn.

- 1) Results from all the participating laboratories are impressively close.
- 2) Agreement of the pooled means, whether weighted or not, demonstrates that interlaboratory alignment can be maintained from modern to at least 20,000 years, when ideally sized samples are available. (In this experiment laboratories were free to choose whatever size samples they required).
- 3) The distribution of the individual results support the one sigma error estimates commonly quoted for the appropriate age levels for averagely sized samples (eg, ~5g equivalent carbon).
viz $\sim\pm 50$ yr for samples up to 5000 years
 $\sim\pm 100$ yr for samples up to 10,000 years
 $\sim\pm 250$ yr for samples up to 20,000 years

In most cases, however, these estimates refer to only part and not the full processing procedure which actual samples undergo.

- 4) The pooled means provide valuable calibration values for the benzene stock solutions now held by the organizing laboratories.

This experiment was intended as a trial run to indicate what might be achieved from a much wider, perhaps even world-wide, application of intercomparison tests. Although stock solutions that allow tests by at least 50 more laboratories are available, such tests would no longer be blind in that results for each level have already been published. However, requests for aliquots of these solutions can be submitted and the organizing laboratories will be interested in receiving results.

With support, it may be worthwhile to use new solutions, in conjunction with the existing ones, if a more general intercomparison is required. In the meantime, we hope to continue the planned program in the United Kingdom, preparing and distributing different sample types at various levels, namely oxalic acid and selected natural samples. Suggestions will, of course, be welcome.

We particularly hope that an important result of this intercomparison will be increased confidence in measurement procedures for users of ^{14}C . We also hope that the knowledge that variation in date results from one laboratory to another need not be significant will be of help in the interpretation of multi-laboratory results.

ACKNOWLEDGMENTS

Thanks are due to the members of staff of the organizing laboratories who have taken part in the measurement work and in particular to D B Punter (Harwell) for his assistance in making up the dilutions. Thanks are also due to the laboratories who have contributed: Radiocarbon Dating Unit, Palaeoecology Laboratory, The Queen's University of Belfast (G Pearson); Radiocarbon Dating Laboratory, Department of Geological Sciences, University of Birmingham (R Williams); The Godwin Laboratory, Sub-Department of Quaternary Research, University of

Cambridge (R Switsur); Radiocarbon Laboratory, Department of Botany, University College, Cardiff (Q Dresser); Natural Environment Research Council Radiocarbon Laboratory, Scottish Universities Research and Reactor Centre, East Kilbride (D Harkness); Radiocarbon Laboratory, Chemistry Department, The University, Glasgow (M Baxter and M Stenhouse).

APPENDIX

EXPLANATION OF DATA UNITS USED IN THE TABLES
(Definitions following Stuiver and Polach, 1977)

1) *Percent modern* (as used for level A)

$$\text{Percent modern} = \frac{A_s}{A_{\text{on}}} \times 100 (\%)$$

where, A_s is the relative activity of the sample not corrected for fractionation.

A_{on} is the corresponding relative activity of 0.95 times the oxalic acid corrected for fractionation ($\delta^{13}\text{C}$).

$$\text{viz, } A_{\text{on}} = 0.95 A_{\text{ox}} \left(1 - \frac{2(19 + \delta^{13}\text{C})}{1000} \right)$$

2) *Equivalent age 'bp'* (as used for levels B to E)

$$t = 8033 \ln \frac{A_{\text{on}}}{A_s} \text{ (years)}$$

(A_{on} and A_s as above)

3) $d^{14}\text{C}$ (given by some laboratories and used to calculate % modern and age 'bp')

$$d^{14}\text{C} = \left(\frac{A_s}{A_{\text{on}}} - 1 \right) \times 1000 (\text{‰})$$

$$\text{then, (i) Percent modern} = \left(1 + \frac{d^{14}\text{C}}{1000} \right) \times 100 (\%)$$

$$\text{(ii) Age 'bp'} = -8033 \ln \left(1 + \frac{d^{14}\text{C}}{1000} \right) \text{ (years)}$$

REFERENCES

- Clark, R M, 1975, A calibration curve for radiocarbon dates: *Antiquity*, v 49, p 251-264.
- Florkowski, T, Payne, B R, and Sanzay, G, 1970, Interlaboratory comparison of analysis of tritium in natural waters: *Internatl Jour Appl Radiation and Isotopes*, v 21, p 453-458.
- Fraser, I, Polach, H A, Temple, R B, and Gillespie, R, 1974, Purity of benzene synthesized for liquid scintillation C-14 dating, in Stanley, P E and Scoggins, B A, eds, *Liquid scintillation counting, recent developments*: New York, Academic Press.
- Otlet, R L and Warchal, R M, 1978, Liquid scintillation counting of low-level ¹⁴C, in Crook, M A and Johnson, P, eds, *Liquid scintillation counting*, vol 5: London, Heyden, p 210-218.
- Polach, H A, 1972, Cross checking of the NBS oxalic acid and secondary laboratory radiocarbon dating standards, in Rafter, T A and Grant-Taylor, T, eds, *Internatl conf on radiocarbon dating*, 8th, Proc: Wellington, Royal Soc New Zealand, v 2, p 688-717.
- 1979, Correlation of ¹⁴C activity of NBS oxalic acid with Arizona 1850 wood and ANU sucrose standards, in Berger, Rainer and Suess, H E, eds, *Radiocarbon dating*, Internatl radiocarbon conf, 9th, Proc: Berkeley/Los Angeles, Univ California Press, p 115-124.
- Stuiver, Minze and Polach, H A, 1977, Discussion: Reporting of ¹⁴C data: *Radiocarbon*, v 19, p 355-363.
- Ward, G K and Wilson, S R, 1978, Procedures for comparing and combining radiocarbon age determinations: A critique: *Archaeometry*, v 20, pt 1, p 19-31.

DISCUSSION

Suess: How do the actually observed errors compare with the theoretical statistical minimal errors? Are they 10, 20 or 50 percent larger or are they larger by a factor of 2, 3, or more?

Otlet: I don't think we are able to give a precise answer to this question. The actually observed errors which we quote in the text are simply calculated from the distribution of the individual means given to us by the laboratories on their returns. As stated, we preferred not to quote their associated error estimates because we felt this would, in a number of cases, disclose the identity of the measurement laboratories and we had promised anonymity. In fact the error estimates varied considerably above and below the average values which was to be expected because of the different measurement procedures. It would take a more detailed analysis to comment on the specific validity of each estimate but, without this, the impression is formed that most people allow for more than simple Poisson counting statistics in their error estimates, and in the returns made for this experiment, there may have been a tendency to overestimate the true experimental errors perhaps by 10 to 20 percent.