

## AN OVERVIEW OF ALL THREE STAGES OF THE INTERNATIONAL RADIOCARBON INTERCOMPARISON

E M SCOTT<sup>1</sup>, T C AITCHISON<sup>1</sup>, D D HARKNESS<sup>2</sup>, G T COOK<sup>3</sup> and M S BAXTER<sup>3</sup>

**ABSTRACT.** The International Collaborative Study involved a wide range of sample materials and ages and, on completion, assessed each stage independently (Scott *et al* 1989; Aitchison *et al* 1990). We combine here the three stages of the study and provide an overview of the uncertainties in the dating procedure as a whole and in its component processes. Three key optimal performance indices, related to internal and external precision and to bias, have been defined to allow quantitative assessment of *Internal Consistency* and *External Consistency* (Aitchison *et al* 1990). We believe that these measures provide quantitative descriptions of a laboratory's reproducibility, accuracy and precision.

For the internal consistency, we have defined the Internal Error Multiplier of the quoted error and, for the external consistency of any laboratory relative to an appropriate baseline, we have defined two indices, the Systematic Bias and External Error Multiplier of the quoted error. We have evaluated the three indices over the three stages and have assessed the relative performances of gas counting, accelerator and liquid scintillation laboratories. The quoted errors describe adequately the variability in duplicate results, but there is evidence of systematic biases and underestimation of interlaboratory variability. We have considered the contribution of pretreatment, synthesis and counting to the overall variability for each laboratory type. We found that, for liquid scintillation (LS) and gas counting (GC) laboratories, ca 66% of the total variation is due to counting and sample synthesis whereas, for accelerator mass spectrometry (AMS) laboratories, the major sources of variability are the sampling and pretreatment processes.

### INTRODUCTION

The International Collaborative Study generated a wealth of data on the analytical variability of the processes involved in radiocarbon dating. It provided an opportunity to justify the quoted error experimentally and to assess the contributions by the component processes of counting, synthesis and pretreatment to the overall variability of dating.

The project, undertaken by a representative subset of the <sup>14</sup>C community, was the largest and most extensive study of laboratory performance in the history of radiocarbon dating. It involved considerable effort on the part of all participants and resulted in a data base of over 600 <sup>14</sup>C dates. The study took 4 years to complete, and involved over 50 laboratories worldwide. Table 1 lists the participant laboratories. The key design features of the study are:

1. *Hierarchical structure.* An additional experimental process was introduced at each of the three stages. Stage 1 concentrated on the counting process, Stage 2 introduced sample synthesis and Stage 3 involved full assay of real samples.

2. *Replication.* Unidentified duplicate samples were included in each stage to allow assessment of internal consistency relative to the quoted errors. The sample materials included calcium carbonate and benzene for Stage 1, humic acid, algal lithothamnion and cellulose for Stage 2, and wood, shell and peat for Stage 3. The age range of material was ~ modern to 7000 BP. Laboratories were issued appropriate sample information and asked to report the results and corresponding errors in standard format. We present here the first analysis of results from all three stages. We address three key topics in particular:

1. The role of the quoted error as a measure of internal consistency as indicated by the duplicate analyses
2. The existence, or otherwise, of systematic biases and the role of the quoted error in adequately explaining any such inter-laboratory variation
3. A comparison of the performance of each laboratory type (*ie*, LS, GC and AMS).

<sup>1</sup>Department of Statistics, Glasgow University, Glasgow G12 8QW, Scotland

<sup>2</sup>NERC Radiocarbon Laboratory, East Kilbride, Glasgow G75 0QU, Scotland

<sup>3</sup>Scottish Universities Research and Reactor Centre, East Kilbride, Glasgow G75 0QU, Scotland

TABLE 1  
Participant laboratories

ANU	-	Radiocarbon Dating Research, ANU, Canberra, Australia
B	-	Physikalisches Institut, Universität Bern, Bern, Switzerland
Deb	-	Institute of Nuclear Research, Debrecen, Hungary
ETH	-	ETH AMS Facility, Zürich, Switzerland
FZ	-	Departemento de Fisica, Fortaleza, Brazil
Gd	-	Institute of Physics, Silesian Technical University, Gliwice, Poland
Gif	-	Centre des Faibles Radioactivités, Gif-sur-Yvette, France
GrN	-	Isotope Physics Laboratory, Groningen, The Netherlands
GSC	-	Radiocarbon Laboratory, Geological Survey of Canada, Ottawa, Canada
GU	-	SURRC <sup>14</sup> C lab, East Kilbride, Glasgow, Scotland
HAM	-	Isotope Dating Laboratory, University of Hamburg, Hamburg, Federal Republic of Germany
Hd	-	Radiocarbon Laboratory, University of Heidelberg, Heidelberg, Federal Republic of Germany
Hel	-	Radiocarbon Laboratory, University of Helsinki, Helsinki, Finland
Hv	-	Niedersächsisches Landesamt für Bodenforschung, Hannover, Federal Republic of Germany
K	-	Radiocarbon Laboratory, National Museum, Copenhagen, Denmark
KI	-	Radiocarbon Laboratory, Kiel, Federal Republic of Germany
KR	-	Institute of Physics, Krakow, Poland
LE	-	Institute of Archaeology, Leningrad, USSR
LP	-	Laboratorio de Tritio y Radiocarbono, La Plata, Argentina
Lu	-	Radiocarbon Dating Laboratory, University of Lund, Lund, Sweden
Lv	-	Radiocarbon Laboratory, University of Louvain, Louvain, Belgium
Ly	-	Laboratoire de Radiocarbone, Lyon, France
NTU	-	Radiocarbon Laboratory, National Taiwan University, Taipei, Taiwan, China
NZ	-	Institute of Nuclear Sciences, Lower Hutt, New Zealand
QL	-	Quaternary Research Center, University of Washington, Seattle, Washington, USA
QLA	-	AMS facility, University of Washington, Seattle, Washington, USA
R	-	Department of Physics, Rome, Italy
RIDDL	-	RIDDL Group, Simon Fraser University, British Columbia, Canada
RT	-	Isotope Department, Weizmann Institute of Science, Rehovot, Israel
SMU	-	Radiocarbon Laboratory, Southern Methodist University, Dallas, Texas, USA
SRR	-	NERC <sup>14</sup> C lab, East Kilbride, Glasgow, Scotland
Su	-	Geological Survey of Finland, Espoo, Finland
T	-	Radiological Dating Laboratory, Trondheim, Norway
Tx	-	Radiocarbon Laboratory, University of Texas, Austin, Texas, USA
U	-	Radiocarbon Laboratory, Uppsala, Sweden
Ua	-	Tandem Accelerator Lab, Uppsala, Sweden
UBAR	-	Dept of Quimica Analitica, University of Barcelona, Barcelona, Spain
UD	-	Centro de Ricerca Applicata, Udine, Italy
UGRA	-	Radiocarbon Laboratory, University of Granada, Granada, Spain
UtC	-	Rijksuniversiteit, Utrecht, The Netherlands
VRI	-	Vienna Radium Institut, Vienna, Austria
WAT	-	Earth Sciences Department, University of Waterloo, Waterloo, Canada
Wk	-	Radiocarbon Laboratory, University of Waikato, Hamilton, New Zealand
Z	-	Rudjer Boskovic Institute, Zagreb, Yugoslavia
	-	Water Resources Research, Adelaide, Australia
	-	ENEA, Bologna, Italy
	-	ORSTOM, Centre de Bondy, Bondy, France
	-	Dept of Physics, Nanjing University, Nanjing, China
	-	Isotope Hydrology Laboratory, Islamabad, Pakistan

In this report, we have concentrated on the results from labs that completed a minimum of 2 of the 3 stages (but including the third and final stage). A total of 38 laboratories satisfied this condition.

DATA DESCRIPTION

Typically, a laboratory that completed all three stages generated a total of 16 radiocarbon dates on 9 distinct samples, 7 assayed in duplicate. As a convenient means of summarizing each laboratory's results, we have calculated two quantities:

1. *Laboratory disparity* calculated on the basis of the duplicate samples. This quantity is simply the unsigned difference between duplicates divided by the square root of the sum of squares of the quoted errors. Values for the disparity of > 1 indicate that the discrepancy between the duplicate samples was greater than expected given the quoted errors. The disparity data form the basis of the assessment of a laboratory's analytical reproducibility (*ie*, precision).

2. *Laboratory offset* is an estimate of the difference between the lab result for a specific sample and its "true value." More typically, here the latter is a robust measure (median) of the consensus value based on all the study results for that sample. The laboratory offset data allows us to assess the extent of interlaboratory variation and the existence of systematic biases.

Figure 1 shows the disparities for each sample assayed in duplicate across all three stages. The majority of disparities are < 1, but they increase in value from stage to stage. At each stage, there are outlying values for which the difference between duplicates is much larger than expected on the basis of the quoted errors. Similar features are apparent if this diagram is reproduced for each of the three different lab types.

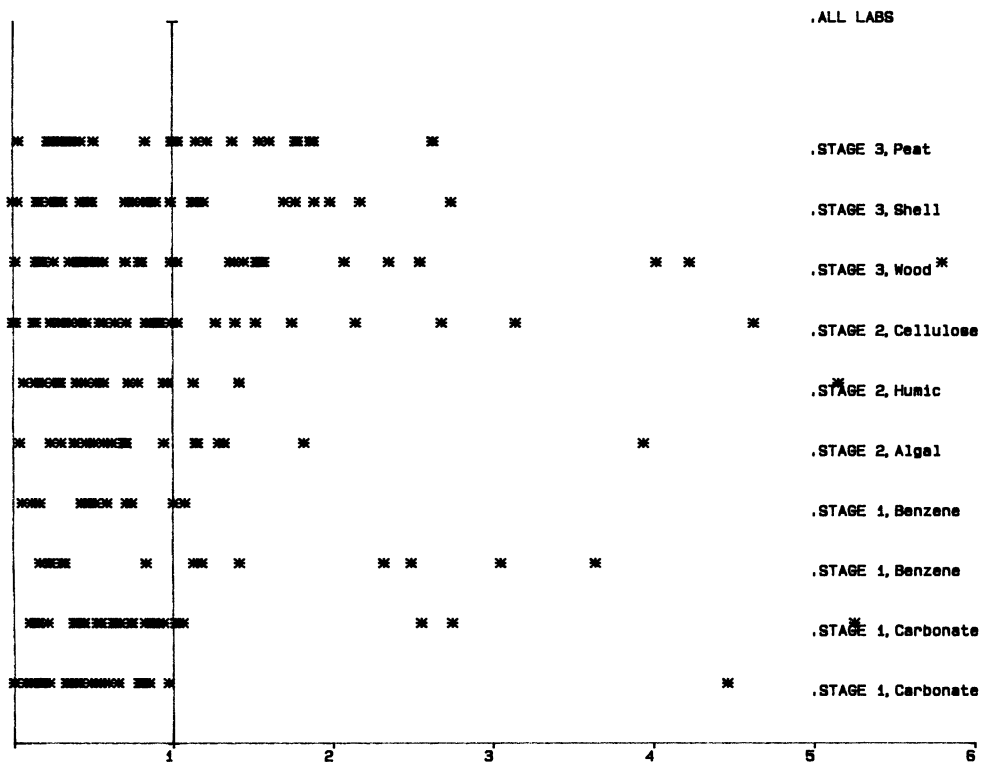


Fig 1. Disparities at each stage for all laboratories

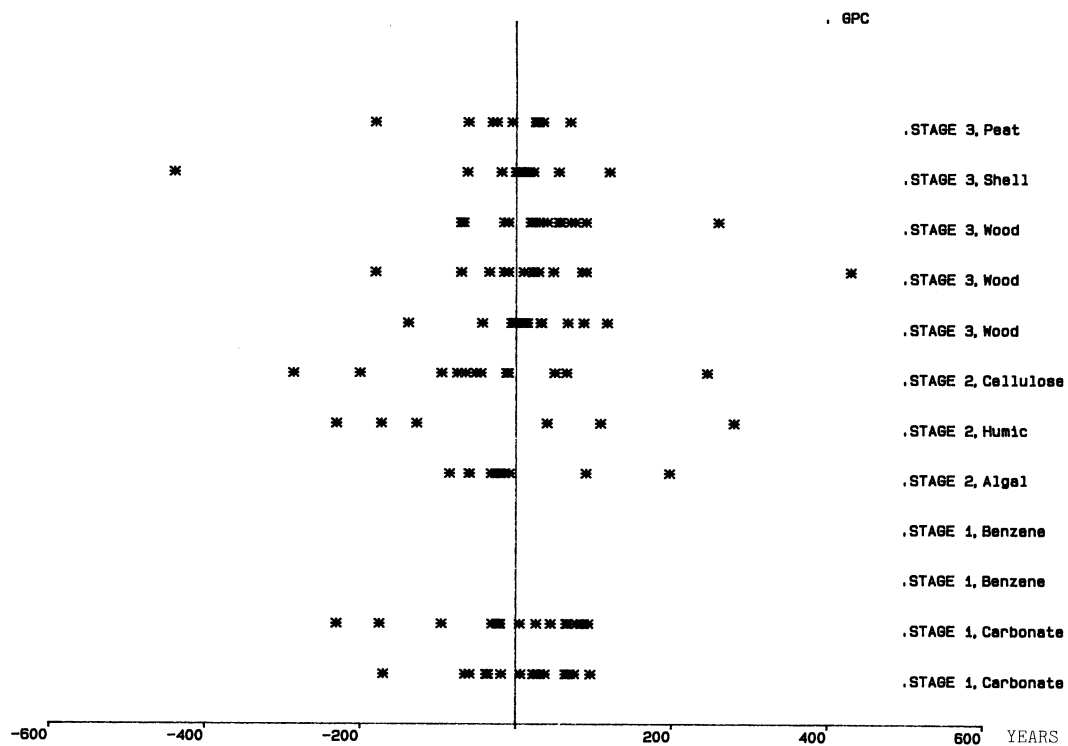


Fig 2A. Offsets at each stage for GC laboratories

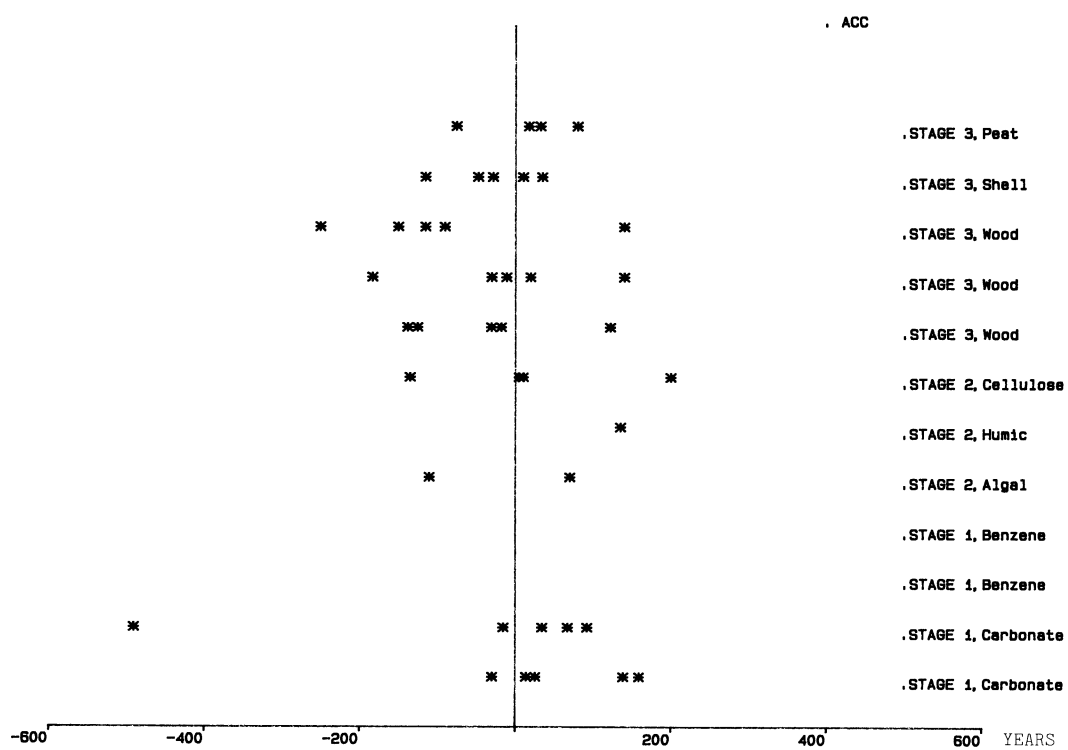


Fig 2B. Offsets at each stage for AMS laboratories

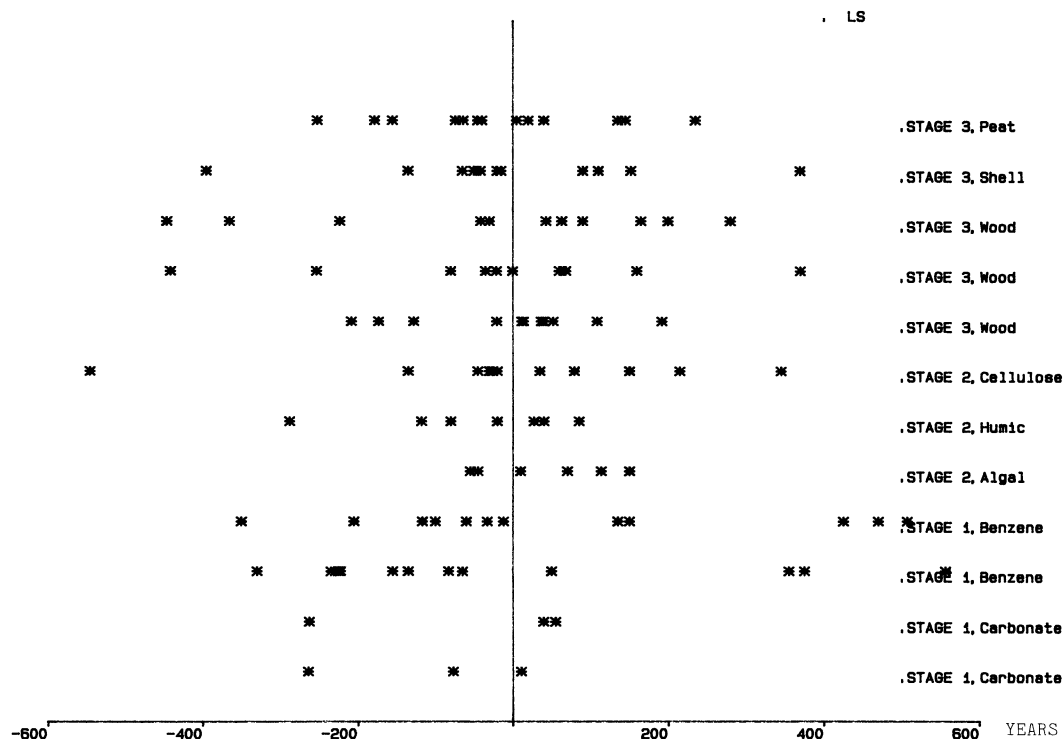


Fig 2C. Offsets at each stage for LS laboratories

Figures 2A, B, C show the laboratory offsets at each stage of the study and for the three distinct lab types. Again, single outliers are apparent at each stage, but for the liquid scintillation laboratories, the variation is larger than that for the gas counting and accelerator labs.

#### DATA ANALYSIS

To quantify this interpretation, we have defined three measures of laboratory performance; two are linked to quoted error through the concept of an error multiplier.

The *Internal Error Multiplier* (IEM) is based on the disparities, and used as a measure of internal precision; the second, the *External Error Multiplier* (EEM), relates quoted error to external variation; the third, *Bias*, quantifies systematic bias. Both EEM and Bias are based on the laboratory offsets. For error multipliers, a value considerably exceeding 1 would indicate that the quoted errors inadequately describe the variation in the laboratory's results.

#### INTERNAL CONSISTENCY

Figure 3 shows the internal error multiplier at each of the three stages, for each of the three laboratory types. The values increase slightly across the stages. Table 2 gives the median values at each stage and confirms this finding. In particular, values of IEM are higher for liquid scintillation labs than for other lab types and AMS labs show a large increase from Stage 2 to 3. However, for individual labs, no clear pattern emerges across the three stages.

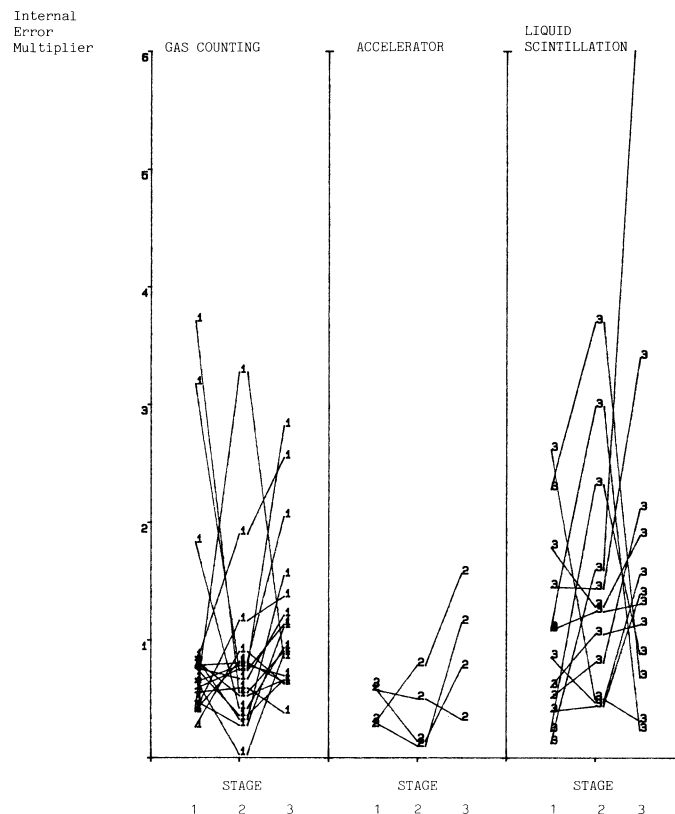


Fig 3. Plots of the IEMs estimates across the 3 stages

TABLE 2  
Internal Error Multiplier  
median

Lab type	Stage 1	Stage 2	Stage 3
LS	1.08	1.28	1.30
GC	0.64	0.75	0.89
AMS	0.31	0.49	1.14

In addition, using the combined results for each laboratory, we have evaluated a plausible range of values for the IEM (formally a 95% confidence interval). We note that, in the absence of a clear pattern across the three stages, the combined IEM is likely to overestimate the multiplier and result in wider confidence intervals. However, if the whole of this interval exceeds 1, then evidently the laboratory is underestimating its internal precision. Figure 4 shows the interval estimate for IEM for each laboratory; the different laboratory types are clearly indicated. Five gas counting and 5 liquid scintillation laboratories have intervals wholly exceeding 1, suggesting that their quoted errors are too small.

#### EXTERNAL CONSISTENCY

We have evaluated systematic bias relative to a baseline defined by all the study results. Figure 5 shows the estimates of systematic bias at each stage. Some laboratories show evidence

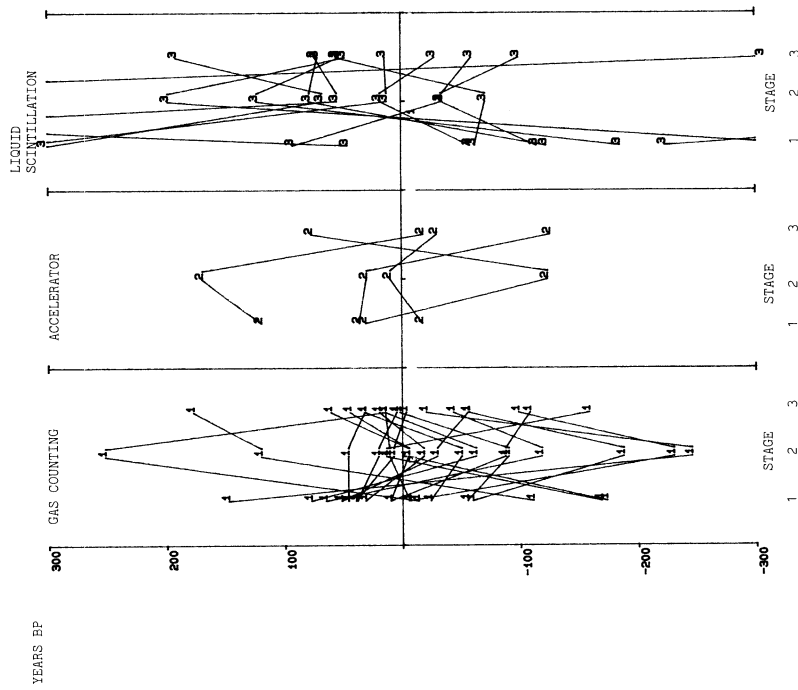


Fig 5. Plots of systematic bias estimates across the 3 stages

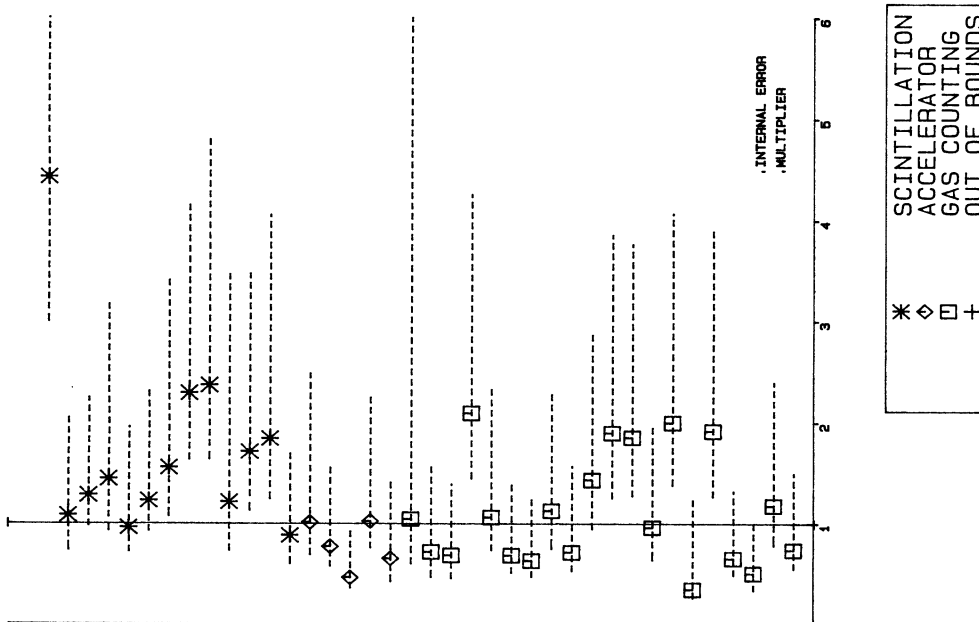


Fig 4. 95% confidence intervals for IEMs

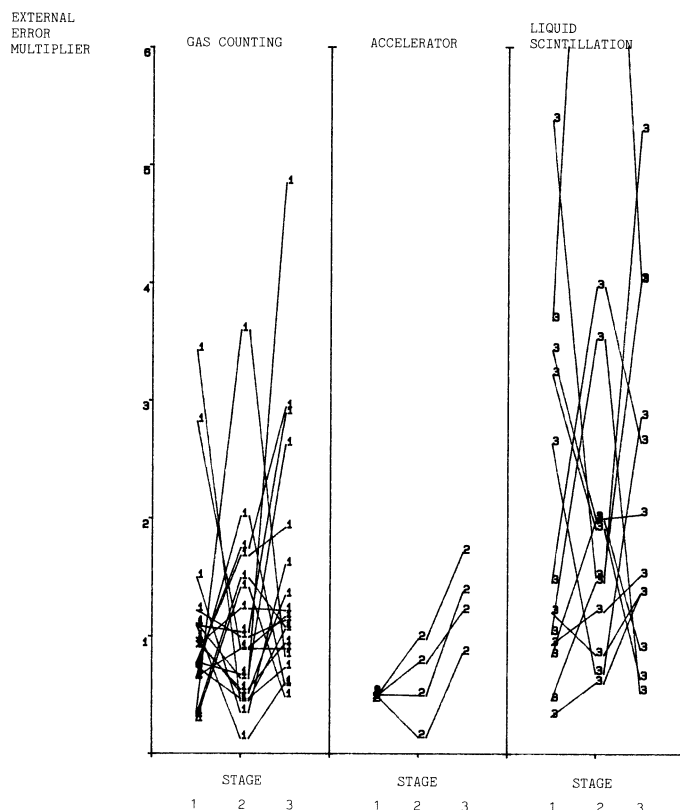


Fig 6. Plots of EEM estimates across the 3 stages

TABLE 3  
External Error Multiplier

Lab type	Stage 1	Stage 2	Stage 3
LS	1.82	1.95	2.41
GC	1.16	1.17	1.42
AMS	0.60	0.95	1.68

that bias is changing on a relatively short-term basis (from stage to stage). Figure 6 similarly shows the external error multipliers (accounting for bias) evaluated at each stage, further evidence of increasing EEM across the three stages. Table 3 shows the median values at each stage, confirming the general trend.

Figure 7 shows a scatter plot of EEM against systematic bias for all laboratories, with the laboratory type clearly indicated. This diagram shows that a small group of laboratories have satisfactorily small biases (*ie*, plausibly 0) and adequate quoted errors: none of these laboratories uses liquid scintillation.

All combined results form a basis for evaluating ranges of plausible values for these two key laboratory parameters. Figure 8 shows the range of values for systematic bias; the intervals that do not include 0 indicate a significant systematic bias. In this case, 5 LS, 1 AMS and 6 GC labs have significant biases of up to several hundred years.



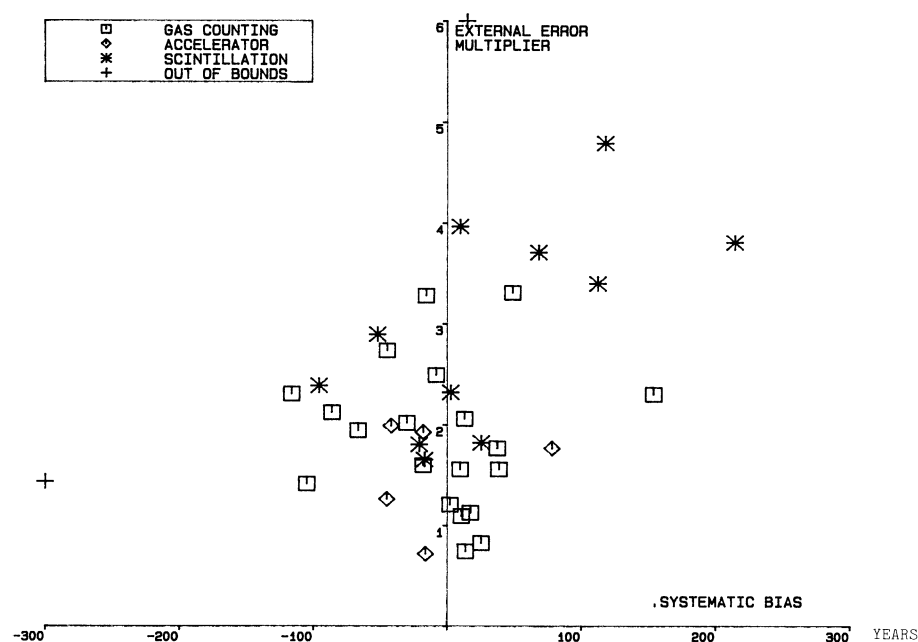


Fig 7. Plot of EEM against systematic bias

Figure 9 presents the equivalent intervals for the external error multiplier for each laboratory, intervals not including 1 indicate that the quoted errors do not compensate for the observed variation around the baseline, even accounting for bias. We point out, however, that laboratories experiencing considerable bias change at each stage (either in sign or magnitude), will also experience amplification of the external error multiple as a result. Nevertheless, no LS laboratory has an EEM that could be 1, only 2 AMS labs and 6 GC labs have EEM that could plausibly be 1, indicating their quoted errors are adequate.

#### OVERALL PERFORMANCE

We have defined three criteria of acceptability, namely no significant systematic bias and adequate assessment of internal and external variability. Table 4 summarizes the assessment of performance from the combined results of all three stages. The major mode of failure is the inadequate assessment of external variability, suggesting that the commonly quoted errors satisfactorily describe the internal variability but not the variation between laboratories.

TABLE 4  
Number of labs failing to meet the criteria

Lab type	IEM	Systematic	EEM
LS	5	5	13
GC	5	6	14
AMS	0	1	3

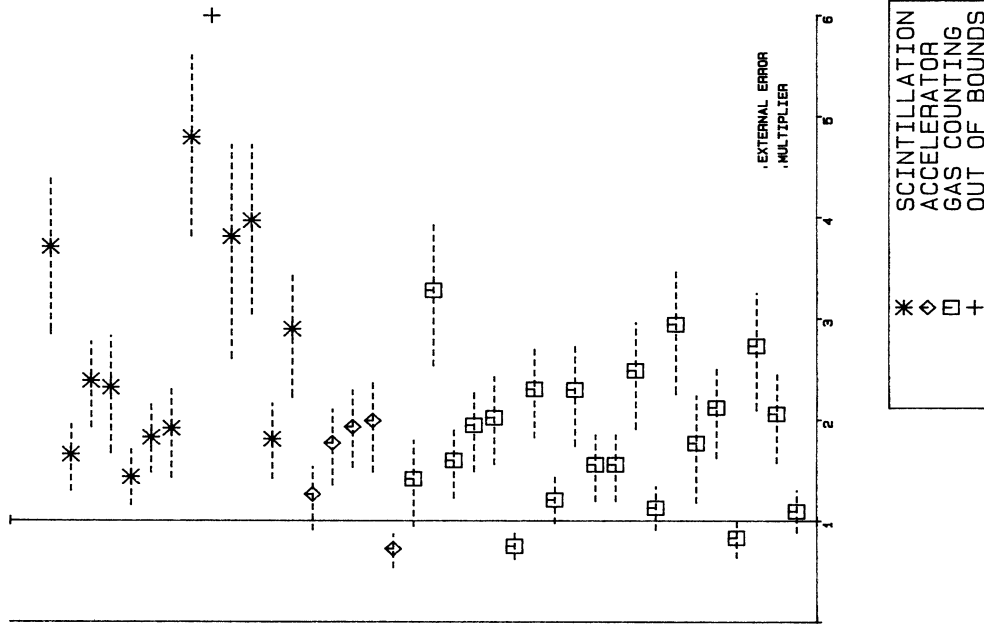


Fig 9. 95% confidence intervals for EEMs.

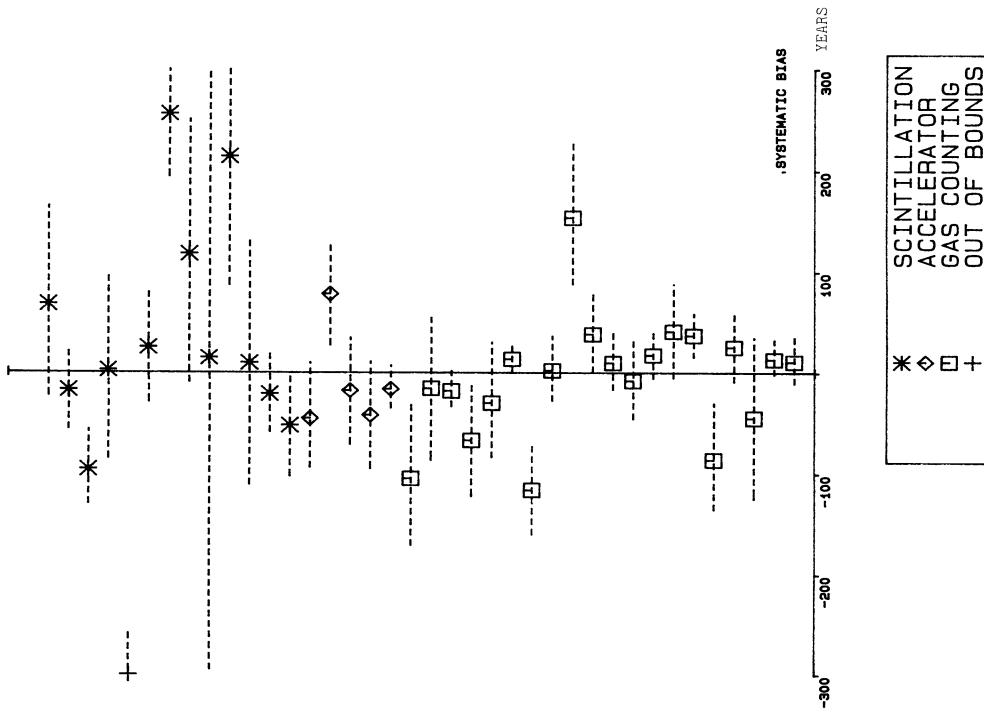


Fig 8. 95% confidence intervals for systematic bias

Provisionally, we can attempt a breakdown of the components of variation across the three stages for each laboratory type. Table 5 shows the percentage of the total variation accounted for in each of the three stages. For both GC, and LS, ca 66% of the variation stems from Stage 1 (primarily counting process for LS labs, counting and synthesis for GC labs). For AMS labs, the major component of variation is introduced in Stage 3, which may be loosely ascribed to sampling and pretreatment.

TABLE 5  
Percentage of total variation

Lab	Stage 1	Stage 2	Stage 3
LS	57	8	35
GC	67	-	33
AMS	13	12	68

#### CONCLUSIONS

At the completion of the largest study undertaken by the radiocarbon community, the findings differ little in nature from those previously reported (ISG 1982).

First, with the inclusion of duplicate samples, we have been able to demonstrate that the quoted errors cover well the internal precision of results. This statement applies to all three stages of the study and for all laboratory types.

Second, we again find evidence of systematic biases amongst the laboratories. The magnitude and sign of the biases have, for some laboratories, changed from stage to stage, indicating relatively short term fluctuations in the source of variation in the results. In some instances, the individual laboratory has identified and corrected the sources of such fluctuations.

Third, we find widespread evidence that the quoted errors do not adequately describe the variation amongst laboratories, even accounting for bias.

The degree of variation observed perhaps reflects difficulties in maintaining suitable and sufficient laboratory standards for calibration against primary standards such as oxalic acid. The type and level of pretreatment applied by individual laboratories varies considerably and, for accelerator laboratories, this, along with the question of representative sampling of the material, is clearly critical. The additional variation in results apparent for liquid scintillation laboratories may well reflect the increasing complexity of current technology.

The radiocarbon community is committed to pursuing these findings and introducing quality assurance proposals to assist each laboratory in maintaining and improving the quality of its results. Users of radiocarbon dates may be assured of the continuation of a program for improvement in what is a complex scientific field.

#### REFERENCES

- Aitchison, TC, Scott, EM, Harkness, DD, Baxter, MS and Cook, GT 1990 Report on Stage 3 of the International Collaborative Program. *Radiocarbon*, this issue.
- ISG 1982 An inter-laboratory comparison of radiocarbon measurements in tree-rings. *Nature* 198: 619-623.
- Scott, EM, Baxter, MS, Harkness, DD, Aitchison, TC and Cook, GT 1989 An interim progress report on Stages 1 and 2 of the International Collaborative Program. In Long, A and Kra, RS, eds, Internatl <sup>14</sup>C conf, 13th, Proc. *Radiocarbon* 31(3): 414-421.