

SECTION 1: THE FOURTH INTERNATIONAL RADIOCARBON INTERCOMPARISON (FIRI)

1.1 INTRODUCTION AND BACKGROUND TO FIRI

Radiocarbon dating is universally used as an essential dating tool in the archaeological and earth (Quaternary) sciences. The technique has enjoyed considerable success with ongoing developments in both the sophistication of experimental practice and an ever-widening range of applications. Most recently (since the 1980s), a new generation of laboratories has been created, based on the exploitation of accelerator mass spectrometry (AMS) for the differentiation and measurement of carbon isotope abundances in natural materials. Worldwide, there are over 100 ^{14}C laboratories now operational in universities, research organizations, museums, and as commercial enterprises. There is an inevitable diversity of experimental approaches and applied priorities within these facilities. Some are well established, while others are relatively recent members of the international ^{14}C community. Consequently, as a group, the laboratories reflect to varying extents the progress achieved over several decades of experience and methodological options. Furthermore, since progress in archaeology and related earth sciences cannot respect geographical and/or present political boundaries, there has been, and continues to be, an inevitable consequence that sample materials from specific cultural contexts are submitted to different laboratories and at different times. In this situation, the issue of comparability between results and amongst laboratories becomes paramount. Users of the results from ^{14}C dating are also concerned with the comparability and quality of laboratory results and the quality assurance programs that laboratories undertake are thus important in ensuring user confidence. The harmonization of measurements and the traceability of results to internationally recognized standards are also major goals of the program of work described in this special issue.

The need for a quantitative assessment of the comparability of ^{14}C dates from diverse laboratories is well recognized and a laboratory intercomparison gives laboratories an opportunity to perform an independent check on their internal quality assurance procedures. In particular, the Fourth International Radiocarbon Intercomparison (FIRI) aimed to define and improve the overall level of confidence in the direct comparison of information obtained from necessarily different measurement systems. This was achieved through a program that focused on assessing and establishing consensus protocols to be applied in the identification, selection, and sub-sampling of materials for subsequent analysis. A large-scale intercomparison then produced direct evidence on precision and accuracy. The intercomparison provided a quantification of some of the main sources of variation associated with the measurements. Finally, as a result of the intercomparison, a further series of ^{14}C reference materials has been created.

An intercomparison is generally considered to be the best scientific tool to determine the current status of laboratory comparability. In the first instance, this approach presents an invaluable opportunity to individual laboratories for checking procedures and results (i.e., it functions at an individual laboratory level as a sound foundation for formal quality control). Most importantly, however, it fosters a harmonization amongst laboratories while simultaneously providing an independent and verifiable measure of interlaboratory comparability for the user. From the statistical analysis of the pattern of results, individual laboratories are able to identify any systematic offsets and to quantify any additional sources of variation (above those already quantified from the intrinsic random nature of radioactive decay). Secondly, the intercomparison also offers vital information in the evaluation of the associated uncertainty statements with each analysis, since all sources of variation are incorporated and through appropriate design, it is possible to quantify the

source and magnitude of different contributors. Finally, dissemination of the outcome of the intercomparison and the highlighting of perceived issues build towards a better understanding between the provider laboratories and the multi-disciplinary user community.

In keeping with the principles of analytical science, ^{14}C laboratories have always been conscious of the importance of accuracy and precision for their reported results, i.e., the ethic of analytical quality control (QC), which in turn is the foundation for the wider concept of quality assurance (QA). This concern for good quality management within the ^{14}C community is exemplified by the care and effort given to establishing and maintaining primary standards and reference materials.

The internationally recognized primary standard is oxalic acid (NBS-OxI and subsequently NBS-OxII [now NIST]). While oxalic acid has all the physical and chemical attributes of a primary standard (e.g., homogeneity, high purity, stability in storage, a constant and known gram molecular weight, etc.), the quantitative and fractionation free recovery of its component carbon (initially as CO_2) has been problematic and the calibration of a material that is compositionally closer to the vast majority of samples submitted for ^{14}C analysis would be advantageous.

International efforts have resulted in the creation of a second tier of materials, so called secondary standards or reference materials. These include materials such as ANU-sucrose and the IAEA quality control reference series (Rozanski et al. 1992). They have 2 main functions: i) for calibration, to demonstrate traceability; and ii) for quality control, to verify the performance of a laboratory.

For quality control purposes, a reference material is commonly a natural material, so that it behaves as similarly as possible to the samples being measured. Therefore, most reference materials must be certified on the basis of measurement by several laboratories, using different methods and using an independent calibration. Certification is only possible when agreement among laboratories performing the measurements can be demonstrated, usually in an interlaboratory comparison.

As early as 1989, Long and Kalin (1990) stressed that it was incumbent on individual ^{14}C laboratories to engage in a formal program of quality assurance (QA), while Polach (1989) noted that the opportunity for internal checking by individual laboratories in routine ^{14}C measurement was hampered by a lack of suitable quality control (QC) and reference materials. Over an extended period of time, the ^{14}C dating community has created a set of reference materials, resulting from a series of voluntary international interlaboratory comparisons. The work in this issue builds on the previous laboratory intercomparisons that have taken place over the last 20 years (ISG 1982; Scott et al. 1991; Rozanski et al. 1992; Gulliksen and Scott 1995), and has created a further series of additional reference materials that are available to the ^{14}C community.

1.2 GENERAL OUTLINE OF THE SPECIAL ISSUE

This special issue is devoted to the design and analysis of FIRI (and to a lesser degree, to the Third International Radiocarbon Comparison—TIRI—see “Part II” section) and has 4 main components:

1. Description of the selection, collection, and preparation of samples for use in the intercomparison.
2. Design and organization of the laboratory intercomparison;
3. Analysis and interpretation of the results;
4. Discussion of the future needs in the community and how the further tier of reference materials created can be used.

1.3 THE PRINCIPLES IMPLICIT IN THE DESIGN AND ORGANIZATION OF A LABORATORY INTERCOMPARISON

1.3.1 Selection of Samples for Reference Materials to be Characterized in an Intercomparison

For quality control purposes, a reference material is commonly a natural material so that it behaves as similarly as possible to the samples being measured. To ensure the widest possible practical advantages, the materials should be representative of routinely dated materials and their ages should span the full range of the applied ^{14}C timescale. These materials are typically certified on the basis of a laboratory intercomparison; therefore, when selecting samples for an intercomparison, their dual purpose must be considered. Given the importance of ^{14}C dating in chronology construction, ideally some of the samples should be independently dated. The most appropriate material for this purpose is dendrochronologically-dated tree-ring sequences, which are already used to underpin the absolute calibration of the conventional ^{14}C timescale back to approximately 9000 yr BP. Also, because of the considerable use of ^{14}C dating within routine archaeological investigations, several samples should be of archaeological significance. Therefore, several samples were included of particular compositional or contextual interest to satisfy the large archaeological user community. Further, given the long history of intercomparisons in the ^{14}C community, it is also important that samples should link any new intercomparison to past studies. In this way, continuity of laboratory performance can be assessed. Available materials in this category include a marine turbidite, Southern Hemisphere wood, and grain. In this way, a catalog of possible samples can be compiled (Section 1.4.3), the samples collected, and a final decision concerning their use made only after appropriate testing.

1.3.2 Preparation and Testing of Material: General Principles

An essential requirement when using a naturally occurring material as the basis for analytical intercomparison is homogeneity. In some instances, this may require that the material be chemically homogenized. However, in the case where the question of interest concerns the variation in the sample results, the raw material can also be provided to allow direct quantification of the natural variation within a typical raw sample and the extent, if any, to which the preselection procedures influence this.

Within FIRI, dendrochronologically-dated wood samples, by their nature, typically required no preparatory treatment, other than cutting into identical (in tree-ring terms) components. For any non-dendro-dated wood, the material was finely chopped and the cellulose component extracted. This was finely ground, homogenized, and pelleted.

For other materials, the degree of preparation varied from a thorough physical mixing (e.g., a turbidite), to grinding and mixing (a whole peat), to complete chemical homogenization (a humic acid extraction from peat). The materials were then packaged and archived.

All bulk materials, after being homogenized, should be checked by replicate analyses on randomly selected aliquots before distribution to the intercomparison participants. Materials should also be tested at different sub-sample sizes (reflecting one of the key differences between AMS and radiometric measurement). In FIRI, all materials were prepared in one batch and a number of randomly selected sub-samples were taken for testing.

1.4 THE PLANNED LABORATORY INTERCOMPARISON (FIRI)

The planned analytical laboratory intercomparison was similar in design to 2 previous ^{14}C studies (TIRI: Gulliksen and Scott 1995, Scott et al. 1998; and IAEA: Rozanski et al. 1992) and included existing reference materials.

1.4.1 Number of Samples and Study Design

The number of samples had to be balanced between the requirements for meaningful statistical analysis of the data and, of course, the practical commitments of the participating laboratories. It was intended that the study should include a degree of replication (with the identity of replicates withheld from the participating laboratories) to allow a direct assessment of within-lab variation (or repeatability). A final figure was agreed of 10 samples to be analyzed within a 1-yr period. A detailed protocol for the reporting of results was also prepared for distribution to all participants. Specific requests to laboratories, including details of how they calculate the error term associated with a ^{14}C age, were made at this time.

1.4.2 Specific Aims and Objectives for FIRI

The fundamental aims and objectives of FIRI can be simply summarized:

- Demonstration of the comparability of routine analyses carried out by both AMS and radiometric laboratories;
- Quantification of the extent of, and sources of, any variation;
- Investigation of the effects of sample size, pretreatment, and precision requirements on the results.

The design structure was rather simple: the intercomparison included core samples, which all laboratories will measure, and optional samples, representing “typical” materials.

The sample selection criteria were relatively simple to express, but more difficult to satisfy due to the quantity of material required. The criteria were the following:

- i. Homogeneity in ^{14}C activity, either as a natural property or artificially induced;
- ii. The samples’ activities should span the activity range from “modern” to “close to background”;
- iii. Some duplicates should be incorporated;
- iv. Some of the samples should form a link to past exercises;
- v. Samples should be available in sufficient quantity to enable excess material to be retained for archiving as new reference materials;
- vi. Most materials should be suitable for measurement by both AMS and radiometric laboratories
- vii. All samples should be natural and several should be dendrochronologically-dated wood.

These criteria translated into samples which included the following: (i) dendrochronologically-dated wood samples with a limited number of rings or drawn from a plateau on the calibration curve; (ii) samples with only a short growing period, and (iii) samples that could be chemically treated and physically homogenized in bulk.

1.4.3 Catalog of Potential Samples for FIRI

A set of potential samples was first identified and collected. The intention after this stage was to identify a subset of these materials that would form the core samples. Any unused samples could be archived or used for optional samples.

1.4.3.1 Dendro-Dated Belfast Wood I

Professor M Baillie of the Queen's University of Belfast Dendrochronology laboratory provided 11.8 kg of dendro-dated wood. The sample was from a Scots pine tree from Garry Bog, Co. Antrim, Northern Ireland with sample identification number of Q7780. The grid reference for the site is C930074, latitude 54°54'N, longitude 6°33'W.

Age: The sample had 40 annual growth rings dating from 3239 BC to 3200 BC.

1.4.3.2 Dendro-Dated Hohenheim Wood

The dendrochronology laboratory of the University of Hohenheim (Dr M Spurk) provided 9.6 kg of dendro-dated oak. The sample identification number was Pettstadt 262.

Age: The sample had 20 annual growth rings dating from 313 BC to 294 BC.

1.4.3.3 New Zealand Kauri Wood

A sub-fossil Kauri wood sample was obtained from Dr A Hogg of the University of Waikato. A previous Kauri wood sample was used by the IAEA (IAEA-C4) (Rozanski et al. 1992), but this sample was not identical. It was expected that this sample would function as a "close to background" sample (Hogg et al. 1995), being at least 40 kyr. Seventy kg of the sample was received as 4 slabs.

Approximate age: close to infinite age with respect to ^{14}C .

1.4.3.4 Russian Wood from Dogee Barrow

A wood sample (part of a log) of approximately 10 kg covering around 190 annual rings from the burial mound of Dogee Barrow, grave 8, (the Tuva king barrows from Scythia) was provided by Dr G Zaitseva of the Institute of the History of Material Culture. The material was excavated in 1998 and was very degraded. Its approximate age was 2300–2400 BP (Sementsov et al. 1998). The sample had not been dendro-dated, and would require careful homogenization since the calibration curve has a steep slope at this time.

Approximate age: less than 1 half-life.

1.4.3.5 Cambridge Modern Wood

A sample of oak (*Quercus robur*) was obtained from Dr R Switsur of the Godwin Institute for Quaternary Research. The tree was planted around AD 1722 and the material corresponding to the period AD 1820–1880 (a relatively flat area on the calibration curve) was removed to provide a sample of 10.4 kg.

Age: modern

1.4.3.6 Belfast Wood II

A further bulk sample of a similar age to Belfast I was also made available. In fact, the second Belfast sample spanned a contiguous set of rings. This sample was used for pretreatment investigation. The sample, which had a finite 40-yr ring span, was again supplied by Prof M Baillie, The Queen's University of Belfast. The sample was 16.3 kg of Scots pine from the Garry Bog, Co Antrim, Northern Ireland, with a sample identification number of Q7780 was provided.

Age: The dendrochronologically determined age span was 3299 BC–3257 BC.

1.4.3.7 Turbidite Carbonate

This sample was supplied by Dr J Thomson, Southampton Oceanography Centre. The sample was mainly coccolith calcite from a single distal turbidite emplaced on the Maderia Abyssal Plain. A remarkable feature of these turbidites is their homogeneity. The basal layers are graded and inhomogeneous, but are overlain by relatively thick deposits. The material used in this study was derived from the middle ungraded deposit and was considered homogeneous. This turbidite was used in the Third International Radiocarbon Intercomparison (TIRI) as the optional Sample K and 25 laboratories had measured it.

Approximate age: 3 half-lives

1.4.3.8 Ellanmore Peat

This sample was again prepared for TIRI (Sample H) but was not used as a core sample. Twenty-seven laboratories chose to analyze it.

It is finely-ground peat from a well-defined stratigraphic section. The Ellanmore peat occurs as an about 50 cm thick horizon intercalated with glacial diamicts and is exposed in a stream bank section of the Reisgill Burn, Ellanmore, Caithness, Scotland (58°18'N, 3°17'W; Natl Grid Ref ND 237 370).

Approximate age: 2 half-lives

1.4.3.9 St Bees Head (Whole Peat) and St Bees Head (Humic Acid)

This sample is from a coastal cliff deposit at St Bees Head in Cumbria, northwestern England (54°29.5'N, 3°37.5'W; NGR NX 9472 1196), which had been exposed by erosion. The apparently well-humified felted peat deposit is approximately 0.5 m thick, and is overlain by several meters of lacustrine material of Holocene age that is largely mineral in nature. Approximately 20 kg of peat were collected and taken back to the laboratory for pretesting. The site was subsequently re-sampled and approximately 30 kg of peat from a slightly different elevation was collected.

Approximate age: 2 half-lives

1.4.3.10 Modern Barley Mash

A modern sample of barley mash provided in the previous exercise (TIRI) was available. Additionally, a new sample was collected from Glengoyne Distillery, just outside Glasgow. The sample comprises a barley grain residue that is left after fermentation and, as such, is a by-product from the manufacture of malt whiskey. The sample represents a single year's growth (1998) and we collected 20 kg of the sample.

Age: modern

1.4.4 Proposed Treatment of Samples Before Testing and Dispatch

Pretesting of the samples was carried out in the Scottish Universities Environmental Research Centre (SUERC) Radiocarbon Laboratory and at the Natural Environment Research Council (NERC) ¹⁴C Laboratory.

1.4.4.1 Kauri Wood

For the FIRI samples, 2 slabs were taken and sliced into sub-samples of approximately 50 g. All participating radiometric laboratories would receive samples of 50 g. For AMS laboratories, 10 of

the 50 g sub-samples, were selected at random and sub-divided into samples of approximately 5 g. Each AMS laboratory received 5 g samples. No pretreatment was carried out on this material. The 2 remaining slabs were sent to the IAEA for archiving.

Both the SUERC and NERC laboratories carried out pretesting of this sample (Table 1.1).

Table 1.1 Pretesting results ($\text{pMC} \pm 1\sigma$) for the Kauri wood sample

Sample	Test 1	Test 2	Test 3	Test 4
Kauri (A and B)	0.15 ± 0.27	0.11 ± 0.26	0.14 ± 0.26	0.12 ± 0.25
	Test 5	Test 6	Test 7	Test 8
	0.11 ± 0.11	0.20 ± 0.11	0.22 ± 0.11	0.30 ± 0.11

A total of 8 analyses were made on this sample and a weighted mean value of 0.2 ± 0.05 pMC was determined. The 8 samples were selected at random from the bags, which were waiting to be sent out to participating laboratories.

1.4.4.2 Turbidite Carbonate

The sample had been stored in a sealed air-tight container since the last intercomparison study (TIRI), to limit interaction with atmospheric CO_2 . The sample was remixed thoroughly before bagging. Approximately 100-g samples would be sent to each radiometric laboratory and 10 g to each AMS laboratory. No pretreatment was carried out on this sample and laboratories were advised not to pretreat it in any way.

Four test analyses were carried out on this sample. Test 4 was an AMS result in which the graphite target was prepared at SUERC and the final measurement was made at the NSF-AMS facility in Arizona. The consensus value for the turbidite sample from the TIRI study is 18,155 BP, while the weighted mean of the 4 results presented here is $18,150 \pm 90$ BP (Table 1.2).

Table 1.2 Pretesting results ($\text{yr BP} \pm 1\sigma$) for turbidite Sample C

Sample	Test 1	Test 2	Test 3	Test 4
Turbidite (C)	$18,305 \pm 180$	$18,010 \pm 180$	$18,220 \pm 165$	$18,050 \pm 190^a$

^aindicates an AMS result

1.4.4.3 Belfast Dendrochronologically-Dated Wood I

The sample was chopped with a clean chisel to give approximately 40-g sub-samples for radiometric laboratories and 4-g sub-samples for AMS laboratories. No further pretreatment was undertaken on this sample. It was felt that there was no need to carry out any pretesting on this sample because of its provenance.

1.4.4.4 Humic Acid

A sub-sample of St Bees peat was pretreated to produce humic acid and humin. Four tests were carried out on this sample and the fractions were dated radiometrically. Table 1.3 shows the results. It can be noted from the results that there were indications of age differences between the humic acid and the humin. The weighted mean value for humic acid age = $11,180 \pm 50$ BP, while for humin, the weighted mean age = $11,500 \pm 115$ BP.

Such differences had already been observed in results from a nearby profile (Doug Harkness, personal communication). Therefore, it was decided not to use the whole peat sample as a core sample because different pretreatments that laboratories employ could lead to an additional source of variation in age. A second peat sample was collected from the site for humic acid extraction to form Sample E.

Table 1.3 Pretesting of humic acid and humin fractions (yr BP $\pm 1\sigma$) from the first St Bees Peat sample

Sample	Test 1	Test 2	Test 3	Test 4
St Bees humic	11,220 \pm 90	11,270 \pm 110	11,190 \pm 100	11,040 \pm 100
St Bees humin	11,600 \pm 200	11,450 \pm 300	11,610 \pm 250	11,350 \pm 210

The pretreatment for humic acid extraction was as follows:

- The sample was digested in 2% KOH at 80 °C to solubilize the humic acid. The sample was then filtered and re-extracted. A total of approximately 150 liters of humic acid solution were extracted from the peat.
- The humic acid was precipitated by adjusting the pH to below 3 by stirred additions of 2 M H₂SO₄. The humic acid was recovered by filtration through glass fiber filter papers and then washed with cold distilled water and dried.
- The humic acid was then re-dissolved in KOH, re-precipitated, washed, and dried to produce a crystalline humic acid material. This was then sub-sampled to 10 g for radiometric laboratories and 1 g for AMS laboratories. (The first peat sample was subsequently archived.)

1.4.4.5 Barley Mash

In the laboratory, the sample was force dried and physically mixed. It was then sub-sampled, giving 50 g for radiometric laboratories and 5 g for AMS laboratories.

This sample was pretested by selecting 2 bags randomly from the sub-samples. The results of the analyses are presented in Table 1.4. The weighted mean value is 110.3 \pm 0.8 pMC.

Table 1.4 Pretesting results (pMC $\pm 1\sigma$) for the barley mash, Samples G and J

Sample	Test 1	Test 2
Barley mash (G and J)	111.0 \pm 0.6	109.5 \pm 0.6

1.4.4.6 Hohenheim Dendrochronologically-Dated Wood

The sample was chopped with a clean chisel to produce pieces of approximately 40 g for radiometric laboratories and 4 g for AMS laboratories. This sample was not pretested because it was of a known age.

1.4.4.7 Belfast Wood II and Cambridge Cellulose, Pretreatment/Preparation

For both samples, fine shavings were produced by planing the wood with a power plane. The samples were then pretreated using a standard acid/alkali/alkali/acid scheme, followed by bleaching with a solution of sodium chlorite in hydrochloric acid. The bleaching process was repeated and the samples washed with distilled water until they were white in color. The samples were dried at 40 °C to leave a white cellulose fraction. The samples were then mixed and sub-sampled to 10 g for

radiometric laboratories and 1 g for AMS laboratories. No pretesting was carried out since these samples were of known age.

The remaining samples which had been identified, although included in the catalog, were not pre-tested since, *a priori*, they did not meet all the selection criteria. They were considered as potential optional samples.

1.4.4.8 Dogee Barrow Wood, Pretreatment/Preparation

This sample was milled, acid/alkali washed, and further physically homogenized and mixed before testing. Homogenization was demanding due to the nature of the calibration curve corresponding to its approximate age.

1.4.4.9 St Bees Head Whole Peat, Pretreatment/Preparation

A bulk peat sample (~30 kg) was cut from a 10-cm-depth increment from within the 0.5 m deposit. The sample was broken up roughly and dried in an oven at approximately 50 °C. Upon drying, the sample was further broken up and then ground to a fine powder. The sample was then thoroughly mixed several times to produce a homogeneous sample.

1.4.4.10 Ellanmore Whole Peat, Pretreatment/Preparation

The peat was air dried at room temperature, ground to a fine powder, and thoroughly mixed to produce a homogeneous sample. This material, as provided, contains about 40% by weight of carbon.

1.4.5 Optional Samples

A series of other materials were also gathered and are described below (their provenance and expected age). Two sets were considered: those for AMS laboratories only (due to the limited quantities) and those suitable for AMS and radiometric laboratories. These samples were not considered potential core samples.

1.4.5.1 Mammoth Tusk (Supplied by Mr S Gulliksen, NUST, Trondheim)

The total weight of the sample was approximately 265 g, (the carbon content was 3.2% of the bone weight). Although the dentin looked very dense and probably would be hard to contaminate to any depth, there were longitudinal cracks that could carry contaminants to attack the “inner” surfaces.

The geological setting of the find was published by Mangerud et al. (1999): *Boreas* (28):46–80. The tusk was found in 1997 by Herbjørn Heggen and Jan Mangerud in a diamicton in a section at the locality of Byzovaya in Pechora, Republic of Komi, in the northern part of Russia.

Approximate age: >5 half-lives.

1.4.5.2 Mammoth Tusk (Supplied by Prof Kh A Arslanov, St Petersburg University, Pechora Project)

This mammoth bone, (Pechora 98-2092), was collected by Valery Astakhov from the right bank of Ute-Yaha river (tributary to More-Yu river). The sample was picked up close to a section of thick aeolian sand covering the local till at 67°12'N, 59°45'E.

Approximate age: close to background.

1.4.5.3 Mammoth Tusk (Supplied by Prof Kh A Arslanov, St Petersburg University, Pechora Project)

A mammoth bone collected by V Makeev from the left bank of the Balyktarkh river, Kotelny Island, Novosibirsk Ostrova.

Approximate age: >2 half-lives.

1.4.5.4 Leather (Supplied by Dr G Zaitseva, Institute of History of Material Culture)

This sample comprises the remains of leather clothes found on a mummified skeleton in the Dogee Barrow (grave 6). Textiles from the same burial have previously been dated. The leather has not undergone any preservation treatment.

As well as these samples, a number of other samples were also available but did not formally form part of FIRI. These included parchment (donated by Asher Jacob, Jerusalem), woollen fabric (Coptic textile), a textile from a late Scythian barrow, as well as other mammoth tusks.

1.5 HOMOGENEITY TESTING

The key sample requirement in an intercomparison is that of homogeneity, which simply means that any sub-sample taken randomly from the bulk material is representative of that material, and that when dealing with trace element assay, that the trace element is uniformly distributed throughout the bulk material. Individual sub-samples should, therefore, have the same properties. Clearly, these are stringent requirements and we describe briefly the testing of samples for inclusion in the intercomparison.

It was decided that not all samples (in particular, the whole wood samples) would require to undergo full homogeneity testing, but all others would be tested. Those samples that were not included in the full homogeneity testing still underwent screening by 2 laboratories (SUERC and NERC).

1.5.1 Test Design

It was agreed that each material would be independently tested in 2 laboratories (where possible a radiometric and accelerator mass spectrometric laboratory), and that a homogeneity testing protocol would accompany each sample. It was agreed that a minimum of 8 analyses was required for homogeneity testing of each core material. Aliquots were labelled “1 of 8”, “2 of 8”, and so on. This same convention was used by the testing laboratory when returning results. The next section details the material testing protocol and the laboratories which dated the materials

1.5.2 Homogeneity Testing Laboratories

Eight laboratories were involved in the homogeneity testing. They are listed below:

Table 1.5 Homogeneity testing laboratories

Name	Method
SUERC ¹⁴ C laboratory, East Kilbride	Liquid scintillation counting (LSC)
¹⁴ C laboratory, University College Dublin	LSC
¹⁴ C laboratory, Weizmann Institute	LSC
NTNU, Trondheim	Gas proportional counting (GPC)
Tandem accelerator laboratory, University of Uppsala	Accelerator mass spectrometry (AMS)
KIK ¹⁴ C laboratory, Brussels	LSC
¹⁴ C laboratory, University of Groningen	AMS
AMS laboratory, University of Aarhus	AMS
NERC ¹⁴ C laboratory	LSC

1.5.3 Sample Testing Protocols

1.5.3.1 *Glengoyne Barley Mash*

Homogeneity testing was to be carried out by radiometric analyses at the SUERC ^{14}C laboratory, East Kilbride, and the ^{14}C laboratory, University College Dublin.

Eight aliquots, each of 100 g, were provided. Each aliquot contained sufficient material for a number of analyses. Instructions included that a sub-sample, representative of the whole, was to be taken and measured and that the samples were not to be pretreated.

1.5.3.2 *Belfast Cellulose*

Homogeneity testing was to be carried out by radiometric analyses at the SUERC ^{14}C laboratory, East Kilbride, and at the ^{14}C laboratory, Weizmann Institute.

Aliquots of 15 g were provided. If necessary, a sub-sample, representative of the whole, should be taken and measured. The samples were not to be pretreated.

1.5.3.3 *Turbidite Carbonate*

Homogeneity testing was to be carried out by radiometric and AMS analyses at the National C-14 laboratory, NUST, Trondheim, and the Tandem accelerator laboratory, University of Uppsala.

For radiometric laboratories, each aliquot contained sufficient material for a single analysis, while for AMS laboratories, each aliquot contained sufficient material for a number of analyses. In this latter case, a sub-sample, representative of the whole, was to be taken and measured. The samples were not to be pretreated.

1.5.3.4 *Cambridge Cellulose*

Homogeneity testing was to be carried out by radiometric analyses at the KIK ^{14}C laboratory, Brussels, and by AMS at the ^{14}C laboratory, University of Groningen.

Aliquots were provided of approximately 13 g for radiometric measurement and 2 g for AMS. If necessary, a sub-sample, representative of the whole, was to be taken and measured. The samples were not to be pretreated.

1.5.3.5 *St Bees Humic acid*

Homogeneity testing was to be carried out by radiometric and AMS analysis at the NERC ^{14}C laboratory, East Kilbride, and the AMS laboratory, University of Aarhus.

For radiometric laboratories, each aliquot (10 g) contained sufficient material for a single analysis, while for AMS laboratories, each aliquot (1 g) contained sufficient material for a number of analyses. In this latter case, a sub-sample, representative of the whole, was to be taken and measured. The samples were not to be pretreated.

1.5.4 Reporting Results

Each laboratory was instructed as follows:

Results for each analysis were to be reported as the measured enrichment relative to the NBS oxalic acid standard ($\delta^{14}\text{C}$), the conventional age BP, and % modern, each with its 1σ error. The $\delta^{13}\text{C}$ should also be reported (if estimated, then this should be noted). For the purposes of homogeneity

testing, we requested that the laboratories measure the samples to as high a precision as reasonably achievable.

1.5.6 Homogeneity Results

1.5.6.1 Pretesting

The prescreening analyses indicated that the samples had ages spanning the timescale required and showed no signs of gross in-homogeneity. Thus, the samples, other than the whole wood, went forward for full homogeneity testing as detailed in the protocols in Section 1.5. The full set of homogeneity testing results are given in Table 1.6.

Table 1.6 Homogeneity results^a

Sample	1	2	3	4	5	6	7	8
Turbidite (radiometric)	18,069 ± 96	18,093 ± 69	18,109 ± 102	18,245 ± 90	18,120 ± 100	18,314 ± 94	18,087 ± 96	18,219 ± 57
$\delta^{13}\text{C}$	1.3	1.2	1.2	1.35	1.48	1.3	1.3	1.3
Turbidite (AMS)	18,745 ± 80	18,555 ± 95	18,655 ± 85	18,505 ± 85	18,510 ± 75	18,765 ± 85	18,655 ± 85	18,500 ± 90
$\delta^{13}\text{C}$	0.9	1.0	1.0	0.9	1.1	0.9	0.7	1.0
Humic (radiometric)	11,855 ± 50	11,855 ± 50	11,870 ± 50	12,020 ± 50	— ^b	11,875 ± 50	11,925 ± 50	11,975 ± 50
$\delta^{13}\text{C}$	-29.0	-28.9	-29.4	-29.4	—	-29.0	-29.0	-29.0
Humic (AMS)	11,790 ± 55	11,857 ± 55	11,925 ± 60	11,875 ± 60	11,940 ± 70	12,005 ± 55	11,895 ± 60	11,910 ± 75
$\delta^{13}\text{C}$	-28.81	-28.79	-28.80	-28.81	-28.82	-28.85	-28.85	-28.87
Barley mash (1) (radiometric)	111.0 ± 0.34	110.5 ± 0.36	110.4 ± 0.36	110.3 ± 0.36	111.4 ± 0.30	111.2 ± 0.27	110.6 ± 0.35	111.1 ± 0.32
$\delta^{13}\text{C}$	-29.2	-29.2	-29.1	-29.3	-29.1	-29.2	-29.2	-29.2
Barley mash (2) (radiometric)	111.1 ± 0.5	111.0 ± 0.5	111.2 ± 0.5	111.0 ± 0.5	110.7 ± 0.5	110.9 ± 0.5	110.9 ± 0.5	110.8 ± 0.5
$\delta^{13}\text{C}$	-28.9	-28.9	-28.9	-28.9	-28.9	-28.9	-28.9	-28.9
Bcellulose (1) (radiometric)	4540 ± 45	4410 ± 45	4450 ± 50	4520 ± 45	4495 ± 50	4475 ± 40	4340 ± 40	4525 ± 40
$\delta^{13}\text{C}$	-23.4	-23.2	-22.6	-22.1	-23.0	-23.0	-23.0	-22.8
Bcellulose (2) (radiometric)	4430 ± 60	4400 ± 70	4480 ± 60	4350 ± 80	4390 ± 70	4510 ± 60	4470 ± 60	4510 ± 60
$\delta^{13}\text{C}$	-23.6	-23.5	-23.5	-23.6	-23.7	-23.6	-23.6	-23.6
Cellulose (AMS)	99.92 ± 0.6	99.30 ± 0.6	99.62 ± 0.6	98.87 ± 0.6	98.52 ± 0.4	97.65 ± 0.55	98.91 ± 0.38	99.06 ± 0.39
$\delta^{13}\text{C}$	-23.2	-24.5	-25.8	-25.1	-25.2	-18.3	-25.5	-25.2
Cellulose (radiometric)	98.36 ± 0.16	98.55 ± 0.19	98.01 ± 0.60	98.31 ± 0.39	98.11 ± 0.95	98.38 ± 0.86	97.94 ± 0.30	98.25 ± 0.25
$\delta^{13}\text{C}$	-23.81	-24.72	-24.14	-25.25	-25.1	-25.02	-25.02	-24.84

^abarley mash and cellulose given in pMC, all others in age (BP)

^bsample lost

1.5.6.2 Analysis of the Results

The results of the homogeneity testing are summarized in Table 1.7, which shows the summary statistics for each series.

Table 1.7 Basic descriptive statistics

Variable	Mean	Median	Standard deviation
Turbidite (R) ^a	18,157 BP	18,114	90
Turbidite (A) ^b	18,611 BP	18,605	109
Humic(R)	11,905 BP	11,875	72
Humic (A)	11,902 BP	11,902	62
Barley (R1)	110.8 pmC	110.8	0.40
Barley (R2)	110.95 pmC	110.95	0.16
Cellulose (A)	99.106 pmC	99.18	0.761
Cellulose (R)	98.239 pmC	98.28	0.205
Bcellulose (R1)	4469 BP	4485	67.6
Bcellulose (R2)	4442 BP	4450	59.2

^aR indicates radiometric

^bA indicates AMS measurement

From the table, it can be seen that the mean age for each sample pairing with the exception of the turbidite is in good agreement. For the turbidite sample, there appears to be a difference, on average, of approximately 500 yr.

The $\delta^{13}\text{C}$ results were also measured by the participating laboratories and showed consistent results.

A test of homogeneity was carried out for each series separately based on the sum of the squared standardized residuals about the mean value. Under the hypothesis that the set is homogeneous, each test statistic should have a χ^2 distribution with $(n-1)$ degrees of freedom, where n is the number of observations. The results of this test are shown in Table 1.8. For all samples, except the humic, the critical value for the test statistic is 14.07 (for humic it is 12.6). It is clear that all individual laboratory sets are homogeneous, with the exception of the Belfast cellulose (R1). The non-homogeneity in this series is likely due to the relatively small errors quoted for each measurement in that set.

Table 1.8 Homogeneity test

Sample	T(R)	T(A)	H(R)	B(R1)	B(2)	C(A)	C(R)	BC(R1)	BC(R2)
Test statistic	7.51	11.9	12.5	8.6	0.7	13.7	4.5	18.3	5.4
Result	H ^a	H	H	H	H	H	H	Non-H ^b	H

^aH: homogeneous

^bnon-H: inhomogeneous

The summary values—mean and median difference, the standard deviation (StDev), standard error (Semean), the minimum (Min), and maximum (Max)—for the differences between the duplicate pairs are shown in Table 1.9

Table 1.9 shows evidence of a significant difference between the 2 sets of turbidite analyses (95% confidence interval for the average difference of -573.6 to -334.9) and the modern cellulose sets (0.212 – 1.523). All others are in agreement within error.

Table 1.9 Descriptive statistics for differences between the matched aliquots

Variable	Mean	Median	StDev	Semean	Min	Max
Turbidite	-454.3 BP	-456.3	142.27	50.4	-676	-260
Humic	8.6 BP	30.0	94.3	35.7	-130	145
Barley	-0.15 pmC	-0.20	0.524	0.185	-0.8	0.7
Cellulose	0.868 pmC	0.890	0.784	0.277	-0.73	1.61
Bcellulose	26.9 BP	12.5	96.8	34.2	-130	170

Reasons for these apparent differences were then sought.

1.5.7 Issues Raised as a Result of the Testing

1.5.7.1 Turbidite Sample

The turbidite sample had been pretreated in one laboratory and not in the other; thus, it became clear that the sample showed a statistically significant difference dependent on the pretreatment. However, it was encouraging that within each laboratory's procedure, there was no evidence of inhomogeneity. It was decided to carry out further analyses on this sample, which are reported in the following. This analysis focused on the small sample properties, and so was solely carried out by AMS.

The total chemistry and measurement background were checked by leaching an Icelandic carbonate (doublespar) before and after the 8 turbidite samples. The results, representing the 2 graphitization reactors used, were $45,490 \pm 980$ BP; $44,455 \pm 950$ BP; and $45,770 \pm 880$ BP; $50,010 \pm 1115$ BP, respectively.

The quality assurance (QA) was also performed by simultaneous preparation and measurement of a reference humic acid sample with a consensus value of 3352 ± 6 BP. The result was 3325 ± 55 BP.

Both QA checks confirmed that the laboratory procedures were stable.

In the 2nd test, 1 large turbidite sample was leached with 0.5 ml HCl. This outermost fraction was dated to $14,290 \pm 135$ BP. One sample was totally leached with 2M HCl and dated to $18,070 \pm 100$ BP. A 3rd sample was leached in 2 separate steps and the results for the 1st and 2nd fraction gave $17,820 \pm 95$ BP and $18,445 \pm 105$ BP, respectively.

In a 3rd test, the 50% inner fraction, as well as the total sample, were analyzed as:

Sample	Inner	Total
1	$18,495 \pm 75$	$18,020 \pm 95$
2	$18,500 \pm 85$	$18,150 \pm 110$

These results are consistent with the former measurements and give consensus values for total sample and inner 50% fractions of $18,073 \pm 58$ BP and $18,579 \pm 26$ BP, respectively.

The turbidite was also analyzed by x-ray diffraction; 80–85% consisted of calcite (aragonite and dolomite were <1%), 10–15% of quartz, and a few % of feldspar.

An attempt to mechanically separate different grain size fraction by vibration and sliding indicated that the material was too finely powdered to make such a partitioning possible in a simple way (normal sample handling).

Thus, it was concluded that the apparent large age differences reflected purely a pretreatment effect, and that consistent application of the no pretreatment instruction would avoid any problem with this sample.

A further detailed study of this sample also revealed the large radon (^{222}Rn) content of the gas released from this sample. Interim counting of one of the samples had indicated that the gas required storage for a minimum of 60 days to reduce the age shift due to radon contamination to less than 30 yr.

1.5.7.2 Modern Cellulose

The most likely cause of the apparent small difference in the results was due to the use of an in-house modern reference material by the AMS laboratory, which was slightly offset relative to the primary modern standard of NIST OXII.

1.5.7.3 Conclusions

Finally, the sets of results from the 2 laboratories for each material were then combined for homogeneity testing. The humic acid, Belfast cellulose, and barley mash were all homogeneous, but the other materials did not comply, as a result of the significant interlaboratory differences that were identified. The dendrochronologically-dated wood sample (modern cellulose) comes from a well-defined and limited time span, so that the lack of homogeneity observed in this sample likely relates to interlaboratory differences, rather than an intrinsic property of the material.

Notwithstanding these difficulties, the results of the homogeneity testing had indicated that all of the samples—when laboratories complied with specific instructions concerning sample handling and pretreatment—could be considered as homogeneous and, thus, suitable for use in the intercomparison. The homogeneity testing had demonstrated some of the difficulties that could be expected in the full exercise, particularly, the importance of the laboratory calibration to the modern primary standard.

1.6 FINAL SAMPLE LIST

A short list of 7 materials was identified as core samples having met the criteria previously identified and is given in Table 1.10. Three sets of duplicate samples were provided blind (Kauri, Belfast wood, and barley mash). The optional sample list then included Cambridge cellulose (Sample K), Dogee Barrow wood (Sample L), St Bees whole peat (Sample M), 3 mammoth tusk samples (Samples N, O, and P), and leather (Sample Q).

Table 1.10 Core sample descriptions

Core sample description	FIRI code	Age/Activity
Kauri wood	A, B	Near background
Marine turbidite	C	~3 half-lives
Belfast dendro-dated wood	D, F	~1 half-life
Humic acid	E	~2 half-lives
Barley mash	G, J	modern
Hohenheim wood	H	<1 half-life
Belfast dendro-dated cellulose	I	~1 half-life

1.6.1 Instructions and Information Sent to All Participating Laboratories

The samples were then sent to laboratories in August/September 1999. A brief description of the pretreatment undertaken in the laboratory before dispatch and instructions concerning sample

pretreatment (if any) before dating were provided to the participating laboratories. The duplicate samples were not identified and laboratories were asked to treat the samples using their routine laboratory procedures.

A total of 120 sets of samples were dispatched.

For Samples A and B, the laboratories were told that these samples should be considered as close to, or beyond, the limit of ^{14}C detection.

For Sample C, laboratories were told that this sample should be fully hydrolyzed and no fractions should be measured. It was emphasized that this sample required no further pretreatment and that pre-etching had been shown to produce small, but significant, age differences. The laboratories were also informed that this sample should be stored in a sealed container.

1.7 ADDITIONAL INVESTIGATIONS

During the planning of the exercise, it was suggested that 2 separate investigations be undertaken relating to laboratory precision and sample size. Specific samples were identified as being appropriate for these exercises which are described below.

1.7.1 Laboratory Precision

It was suggested that Sample D should be dated to the highest precision possible within the laboratory. We recommended that cellulose be extracted for this study.

1.7.2 Sample Size

Typically, in intercomparisons, samples are provided in sufficient quantity and well-preserved; however, this does not always reflect the “real-life” laboratory situation. Therefore, it was suggested that Sample E, which is approximately 40% carbon, should be dated (where possible) using different sample sizes, and that the sample size (mass of carbon) should be clearly stated when the results were reported. It was suggested that laboratories report the result for Sample E based on their “optimal” sample size and one further result for a sub-sample of the raw material, which was at the minimum sample size for the laboratory to report a meaningful result.

This exercise would allow us to address the question of sub-sampling and the effect of sample size on the final result.

Both investigations were optional, but we believed they were valuable in providing important information.

1.7.3 Result Reporting

Report forms were included, detailing the information requested when returning results. Additional information could be added on separate sheets, if required. The deadline for the return of results was 31 August 2000, but this was subsequently extended to December 2000.