Radiocarbon, Vol 57, Nr 5, 2015, p 1041–1047 DOI: 10.2458/azu_rc.57.18569

ACCURACY AND REPRODUCIBILITY OF ^{14}C MEASUREMENTS AT THE LEIBNIZ-LABOR, KIEL: A FIRST RESPONSE TO LULL ET AL., "WHEN ^{14}C DATES FALL BEYOND THE LIMITS OF UNCERTAINTY: AN ASSESSMENT OF ANOMALIES IN WESTERN MEDITERRANEAN BRONZE AGE ^{14}C SERIES"

John Meadows[1,2] • Matthias Hüls[2,3] • Ralph Schneider[2]

RESPONSE

In this issue, Lull et al. (2015) have compiled a list of anomalous radiocarbon dates on samples of archaeological material from La Bastida and other sites in Spain, reported by our laboratory in 2009–2013. On this basis, they suggest that "at least from 2009 onwards" all Kiel's dates on organic samples are potentially inaccurate, and often too old. We would like to briefly respond to these claims, and to reassure readers that the quality of accelerator mass spectrometry (AMS) ^{14}C measurements at Kiel is comparable again to that in other research laboratories.

We must first acknowledge that Roberto Risch, the submitter of most of these samples, has worked closely with the Leibniz-Labor over several years to try to resolve these issues. Indeed, it is partly thanks to Prof Risch that we became aware of a pattern of increasingly unacceptable results. Known-age standards, e.g. IAEA C1–C7 (C1–C2: carbonates; C3–C7: different organic materials), which were regularly measured together with unknown samples, started to show an increasing scatter around expected ages in the second half of 2010, a pattern that intensified in early 2011 and that appears to have largely been resolved by early 2012 (Figure 1).[4] The organic IAEA standards are sealed for closed-tube combustion to CO_2 without chemical pretreatment, and since inorganic (carbonate) standards, which are converted to CO_2 by acidification, were not affected, a potential problem may have been related to the sealing and combustion steps. Whenever possible, therefore, organic "unknown" samples measured in late 2010 and 2011 were redated from April 2012 onwards, usually in duplicate ("double-sealings," i.e. two targets made from independent combustions of two aliquots of the same extract), and many of the 2010–2011 results were shown to have been inaccurate.[5]

The large number of IAEA standards dated throughout 2012 and 2013 gave acceptable results, and double-sealings generally gave consistent results, but it became clear that these quality-assurance (QA) measures were insufficient to detect all potential problems. Some 2012–2013 results on unknowns that met all acceptance criteria [e.g. carbon yield after pretreatment (%C), visual graphite appearance, ion beam current, δ^{13}C] could still be shown to be unreliable—for example, if

1. Centre for Baltic and Scandinavian Archaeology, Stiftung Schleswig-Holsteinische Landesmuseen Schloss Gottorf, Schleswig, Germany.
2. Leibniz Laboratory for Radiometric Dating and Isotope Research, Christian-Albrechts University Kiel, Germany.
3. Corresponding author. Email: mhuels@leibniz.uni-kiel.de.
4. We do not have a continuous series of measurements on known-age archaeological samples between the end of the Fifth International Radiocarbon Intercomparison (VIRI) in 2008 (Scott et al. 2007, 2010a,b) and the start of 2014. Results from 2014 onwards standards and known-age samples are reported on our web site, http://www.uni-kiel.de/leibniz/Leibniz-web_englisch/qualitytests.htm. Kiel results in the Sixth International Radiocarbon Intercomparison (SIRI) will be posted when consensus values for the SIRI samples are published.
5. Carbonate samples and graphite targets prepared in other laboratories but measured in Kiel (and thus given KIA- numbers) do not appear to have been affected. Because KIA- numbers are assigned to unique samples, not individual measurements, we cannot simply list the KIA- numbers of the organic samples affected (cf. Bronk Ramsey et al. 2004), but relevant submitters have been contacted to arrange redating of organic samples dated in late 2010 and 2011.

measurements from double-sealings were inconsistent.[6] Moreover, this former QA regime was not designed to detect problems during sample pretreatment. In early 2014, therefore, we began a strict "batch" QA system. In essence, unknowns are pretreated concurrently with known-age samples of similar materials and expected date, and this batch of samples passes through every subsequent stage (weighing, sealing, combustion, graphitization and AMS measurement) at the same time. If the AMS results from the known-age samples are unacceptable, the results from unknowns are also rejected, and the entire batch is prepared again from fresh material.

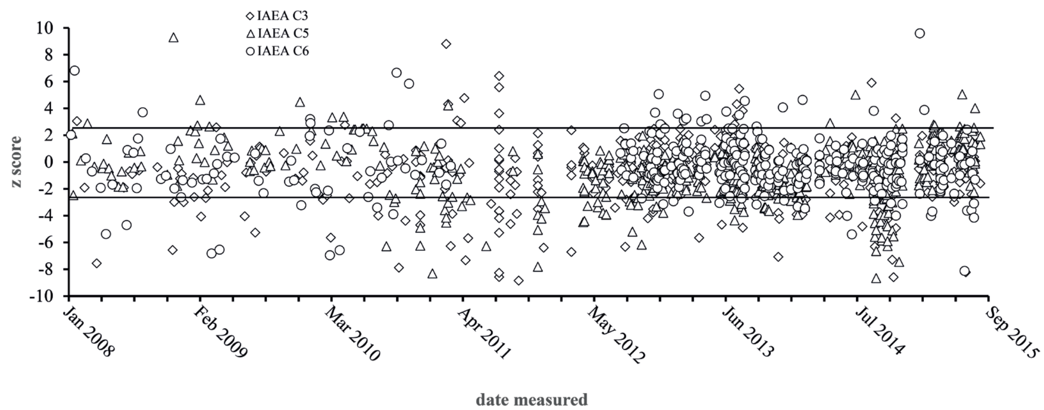

Figure 1 Differences between measured ^{14}C and consensus values, expressed as significance, for IAEA C3, IAEA C5, and IAEA C6 from 2008 to 2015; 95% of measurements should fall within ±2σ of consensus values.

Thirteen of the La Bastida samples first dated in Kiel in 2009 have now been redated under the 2014 batch system: seven samples of bones and six carbonized plant remains. Table 1 lists the 12 samples whose 2014 results are consistent with expectations, based on the La Bastida stratigraphy. KIA-39737, not listed here, gave unexpected results (see below). The same pretreatment methods [i.e. collagen extraction using a modified Longin (1971) method and an acid-alkali-acid (AAA) extraction for organics (Grootes et al. 2004)] were used where applicable, and similar and satisfactory yields were obtained, in 2009 and 2014.[7] As Lull et al. (2015) have noted, collagen preservation in bones from La Bastida is extremely variable, but in these seven bones the collagen yield was at least 3%. The 2014 extractions were either from the same bone sampled in 2009 or from another bone of the same skeleton, in the case of some articulated burials. The plant samples were mainly from deposits of charred grain, and we have no reason to question the excavators' interpretation that the grains in these deposits were uniform in date.

The 2014 results are from batches that included standards prepared and measured concurrently, and which gave acceptable results (Figure 2). We are therefore as confident as we can be that results obtained in 2014 for these 13 unknowns are accurate. Double-sealings from four collagen extracts were dated (KIA-50634, KIA-50636, KIA-50638, and KIA-45106), and the paired results are all statistically consistent. Double-sealings (combustions), with double-graphitizations of sample

6. To gauge whether two or more results are statistically consistent with a single ^{14}C age, throughout this discussion we have used OxCal's R_Combine function (Bronk Ramsey 1995), which applies Ward and Wilson's (1978) test at the 5% significance level. A test statistic, T, is calculated, which should be less than a critical value taken from the χ^2 distribution for the specified significance level and degrees of freedom ($\nu = n-1$) if the results are statistically consistent. If only two results are compared, the Ward and Wilson test is equivalent to a rule-of-thumb that regards two results as inconsistent if the difference between them is >2σ, where σ is the square root of the sum of the squared uncertainties in the individual measurements.
7. With the exception of KIA-39251 (2009)/KIA-50633 (2014), whose 2009 yield was more than double that of the 2014 extraction.

Table 1 Comparison of 2014 measurements of La Bastida samples measured in wheels 1526, 1536, 1543, 1554, and 1562 with pre-2014 measurements. Given are the customer and laboratory ID, the dated fraction (AR: alkali residue; HA: humic acid; AIOM: acid-insoluble organic matter; Lcol: collagen after Longin extraction), single ages, and weighted mean ages where appropriate, and differences as well as significance, either between weighted mean ages (+) or single measurements (*).

	2014 results				Earlier results					
Material	KIA- (Yr-No)	Fraction	Age BP	Weighted mean	KIA- (Yr-No)	Fraction	Age BP	Weighted mean	Difference	Significance
Human bone	50633 (2015) (39251)	Lcol	3298 ± 27		39251 (2009)	Lcol	3408 ± 24		–110 ± 36	–3.05*
Human bone	50634 (2015-1) (39255)	Lcol	3593 ± 26	3604 ± 20	39255 (2009)	Lcol	3523 ± 28		81 ± 34	2.35+
	50634 (2015-2) (39255)	Lcol	3620 ± 32							
Human bone	50635 (2015) (39258)	Lcol	3478 ± 26		39258 (2009)	Lcol	3475 ± 25		3 ± 36	0.08+
Human bone	50636 (2015-1) (39261)	Lcol	3420 ± 26	3425 ± 18	39261 (2009)	Lcol	3354 ± 26		71 ± 32	2.25+
	50636 (2015-2) (39261)	Lcol	3430 ± 25							
Human bone	50637 (2015) (40013)	Lcol	3368 ± 25		40013 (2009)	Lcol	3402 ± 24		–34 ± 35	–0.98*
Faunal bone	50638 (2015-1) (40017)	Lcol	3670 ± 28	3660 ± 22	40017 (2009)	Lcol	3647 ± 29		13 ± 36	0.36+
	50638 (2015-2) (40017)	Lcol	3645 ± 34		40017 (2013)	Lcol	3827 ± 29		–167 ± 36	–4.59+
Faunal bone	45106 (2014-1)	Lcol	3693 ± 32	3690 ± 24	40020 (2009)	Lcol	3640 ± 30		50 ± 38	1.3+
					45106 (2011)	Lcol, So	3592 ± 24		98 ± 34	2.89+
	45106 (2014-2)		3686 ± 35		45106 (2013)	Lcol	3693 ± 30		–3 ± 38	–0.08+
Barley grains	40100 (2014-1.1)	AIOM	3421 ± 31	3423 ± 14	40100 (2009)	AR	3410 ± 62	3399 ± 21	24 ± 25	0.95+
	40100 (2014-1.2)		3464 ± 27							
	40100 (2014-2.1)		3418 ± 31		40100 (2009)	HA	3398 ± 22			
	40100 (2014-2.2)		3390 ± 26							
Barley grains	40101 (2014-1.1)	AR	3376 ± 29	3369 ± 14	40101 (2009)	AR	3353 ± 26		16 ± 30	0.54+
	40101 (2014-1.2)		3328 ± 26							
	40101 (2014-2.1)		3379 ± 29							
	40101 (2014-2.2)		3400 ± 27							
Barley grains	40097 (2014)	AR	3668 ± 33		40097 (2009)	AR	3524 ± 35	3508 ± 21	160 ± 39	4.11+
					40097 (2011)	HA	3499 ± 26			
Barley grains	40098 (2014-1.1)	AR	3601 ± 32	3604 ± 15	40098 (2009)	AR	3455 ± 31	3483 ± 22	121 ± 27	4.54+
	40098 (2014-1.2)		3594 ± 29							
	40098 (2014-2.1)		3622 ± 29		40098 (2011)	HA	3510 ± 30			
	40098 (2014-2.2)		3601 ± 27							
Barley grains	40099 (2014-1.1)	AIOM	3397 ± 29	3398 ± 14	40099 (2009)	AR	3338 ± 162	3402 ± 19	–4 ± 24	–0.17+
	40099 (2014-1.2)		3420 ± 27							
	40099 (2014-2.1)		3397 ± 28		40099 (2009)	HA	3403 ± 19			
	40099 (2014-2.2)		3377 ± 29							

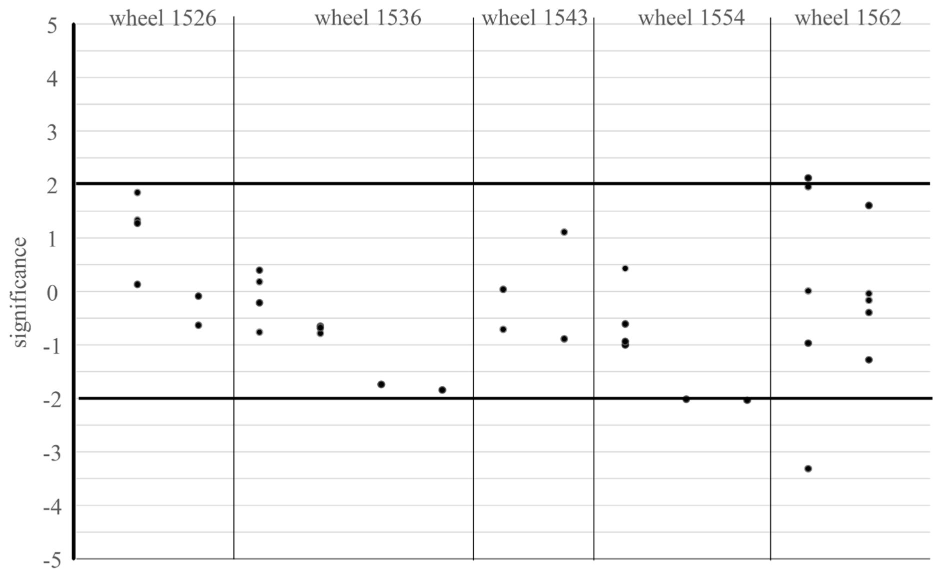

Figure 2 Differences between measured ^{14}C and consensus values, expressed as significance, for relevant standards in wheels 1526, 1536, 1543, 1554, and 1562.

CO_2, were dated from each of two separate extractions of four samples of carbonized grains (KIA-40098–40101), and all gave internally consistent results. KIA-39737 (not listed in Table 1) was not prepared or measured concurrently with the other six plant samples listed and it was the only sample whose 2014 double-sealing dates were statistically inconsistent ($T = 4.3$, $T'(5\%) = 3.8$, $\nu = 1$; Ward and Wilson 1978). Although the standards prepared and measured concurrently gave acceptable results, the 2014 results are both significantly older than expected and inconsistent with 2009 and 2013 measurements for KIA-39737.

Of the other 12 samples, in seven cases (four bones and three plant remains), the 2009 results are statistically consistent with the 2014 dates, and should therefore be accurate. Two of the bones gave slightly younger dates in 2009 than in 2014 (by 71 ± 32 and 81 ± 35 yr, respectively, or 2.25 and 2.35σ), and one bone gave an older date (by 110 ± 36 yr, or 3.05σ). In the last case, the large difference in collagen yields may be a factor in the discrepancy between the 2009 and 2014 results, KIA-39251 and KIA-50633. Finally, two grain samples gave younger dates in 2009 than in 2014 (KIA-40097 by 160 ± 39 yr, and KIA-40098 by 121 ± 27 yr, or 4.1 and 4.5σ). Considering the reproducibility of the 2014 results, we accept that in these two cases the 2009 results were unreliable.

One of these samples, KIA-40020/45106 (Table 1), was also submitted by R Risch to the Mannheim laboratory, and gave a result consistent with our 2009 and 2014 measurements (Lull et al. 2015:Table 6). An extraction in Kiel during the 2010–11 problematic period gave a slightly younger date, and a second extraction in 2013 gave a result consistent with the 2009, 2014, and Mannheim dates. KIA-40017 (=KIA-50638) gave consistent results in 2009 and 2014, and an anomalously old result in 2013, which came from collagen sealed on a date for which most other sealings have been found to be problematic.

Otherwise, the results discussed by Lull et al. (2015) are mainly from samples that have yet not been replicated under the 2014 batch system of quality control. We will briefly comment on them from the laboratory's perspective:

- Burial 23 (Lull et al. 2015:Table 1): both 2009 results for 23/1 (KIA-40014) are clearly too old. The February 2010 result and a second sealing of the 2010 extract, measured in 2013, are consistent with a Mannheim result, and approximately the same date as three results on two extracts of KIA-40015, from the earlier burial in this jar (measured in 2009, early 2010, and 2013). KIA-40014 had a low but acceptable collagen yield (3.5%). The anomalously old results for KIA-40014 in 2009 are still being explored by redating.
- Burial 18 (Lull et al. 2015:Table 1): the collagen yields from both burials in this jar were extremely poor (KIA-45101 ~0.3%; KIA-40012 ~1.4%) and any ^{14}C measurements from "collagen" should be disregarded. The dates on plant remains from the same context (KIA-40099–40100) are perfectly consistent (Table 1).
- Tiera del Lienzo *Pistacia* seeds (Lull et al. 2015:Table 2): the 2010 results for KIA-43101 are clearly too old, but they belong to the problematic period in Kiel, although we were not aware of this at that time. Our 2013 extract gave a result consistent with those obtained by the Mannheim and Uruguay laboratories. According to Lull et al. (2015), additional samples of this material, pretreated in the Leibniz Laboratory during the problematic period but combusted and dated in Mannheim, gave acceptable results.
- Tiera del Lienzo *Hordeum* and *Triticum* grains (Lull et al. 2015:Table 3): the discrepancy between the 2013 Kiel result and the Mannheim result is still unexplained. A new extraction of this sample under the 2014 QA system is currently being redated.
- Cerro de las Palas cereal grains (Lull et al. 2015:Table 4): the 2010 result for KIA-42808 marks the start of the problematic period, and should be disregarded. The 2013 result is significantly different to the Mannheim result, and this sample is also being repeated again under the 2014 QA system.
- Minorcan human remains (Lull et al. 2015:Table 5): these five samples were prepared in Kiel in early 2011, during the problematic period. KIA-44840–44844 are measurements from collagen combusted and graphitized in the Leibniz Laboratory, whereas KIA-46767–46771, deemed satisfactory by Lull et al., are from a second extraction, combusted and graphitized by the KIK-IRPA laboratory in Brussels, but measured in the Leibniz Laboratory. These differences confirm the impression, based on the results of combusted IAEA standards, that a significant number of results during the problematic period were compromised during sealing, combustion, or graphitization.

Lull et al. (2015:Table 6) list 13 other samples submitted by Prof Risch in 2009–2013, with multiple Kiel ^{14}C ages reported as valid under the prevailing quality control systems. The authors state that six redatings gave consistent results, six showed discrepancies, and one was ambiguous. We observe that for at least six of the eight samples submitted in 2009, the original results appear to have been accurate:

- KIA-40017 (=KIA-50638) and KIA-40020 (=KIA-45106) were also redated in 2014 (Table 1), and have already been discussed.
- The 2009 and 2013 results for KIA-40018 (=KIA-45103) and KIA-40019 (=KIA-45105) are statistically consistent; in the second case, Risch also obtained a consistent result from the Mannheim laboratory.
- All results for KIA-40021 (=KIA-45107) are consistent with each other and with a date from Mannheim on the same sample, except for one anomalously young result from a 2011 extract sealed during the problematic period; a second aliquot of the same extract, sealed in September

2012, gave what appears to be an accurate result.

- The 2009 and 2013 results for KIA-40016 (=KIA-45102) are both plausible, but statistically inconsistent with each other and with two unreported measurements during the problematic period; the dating of this sample is therefore unsatisfactory.
- The 2009 result for KIA-40123 is slightly older than the other three measurements—on a 2011 sealing of the 2009 extract, and on sealings of the 2011 extract in late 2011 and late 2012; again, it would be preferable to repeat this sample under the 2014 batch system of quality control.

The other five samples, which as far as we know have not been dated by any other laboratory, were initially dated at the start of the problematic period:

- Results from the 2012 double-sealings of KIA-42493, KIA-42496, and KIA-43166 are consistent with each other, whereas the late-2010 and 2011 results are too old.
- Results from the 2012 double-sealings of KIA-42492 are consistent with each other, and with three of five measurements of this sample during the problematic period.
- The collagen yield from KIA-42495 was <1%, and even the 2013 result should not be regarded as reliable.

To sum up, most of the anomalous KIA- results highlighted by Lull et al. (2015) belong to a period in later 2010 and 2011 in which the laboratory regularly obtained unacceptable results on combusted IAEA standards. Dates from combusted unknowns in this period are thus assumed to be unreliable, and where possible new preparations from these samples have been dated from 2012 onwards. The redating in 2014 of 13 of the 2009 samples, using the more rigorous batch system of quality control, confirms that most 2009 results should be accurate. A small number of 2009 results on well-preserved samples were anomalous, but not consistently too old or too young.

A complete assessment of the ^{14}C results from La Bastida—integrating the full set of dates with stratigraphic constraints on the calendar ages of samples—cannot be presented here, not least because repeat measurements are still ongoing at the Leibniz Laboratory and the submitters have not published the full stratigraphic sequence or provided all the ^{14}C results from other laboratories. Equally, we cannot yet provide an overall assessment of the accuracy or reproducibility of results from unknowns throughout the period 2009–2013 (e.g. if samples were first dated before mid-2010 or after 2011, we would only have repeated them if submitters regarded the results as anomalous). Given the evidence that most problems occurred during sealing or combustion, the pattern of results from combusted IAEA standards (Figure 1) suggests that most results from unknowns dated before the middle of 2010 and/or after the start of 2012 should be reliable. Although not foolproof, the practice of dating double sealings whenever possible provides additional confidence that results reported to submitters in 2012–2013 were accurate. It is clear that the 2014 batch system is more successful at detecting anomalous results before they are reported, but it will take some time before we can quantify the improvement in the accuracy and reproducibility of our results.

ACKNOWLEDGMENTS

The efforts of all Leibniz Laboratory staff over 2009–2014 are gratefully acknowledged. Responsibility for quality control in this challenging period was shared by the writers and Marie-Josée Nadeau, Andrzej Rakowski, Anke Rieck, and Alexander Dreves. Pieter Grootes liaised with the Barcelona team on behalf of the laboratory for most of this period. We are grateful to Roberto Risch and colleagues for their patience and understanding, for providing additional sample material and sharing ^{14}C results from other laboratories in advance of publication. We also thank Mark McClure, managing editor of *Radiocarbon*, for the opportunity to comment on and respond to the paper by Lull et al. in this issue.

REFERENCES

Bronk Ramsey C. 1995. Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* 37(2):425–30.

Bronk Ramsey C, Higham T, Bowles A, Hedges R. 2004. Improvements to the pretreatment of bone at Oxford. *Radiocarbon* 46(1):155–63

Grootes PM, Nadeau M-J, Rieck A. 2004. ^{14}C AMS at the Leibniz-Labor: radiocarbon dating and isotope research. *Nuclear Instruments and Methods in Physics Research B* 223–224:55–61.

Longin R. 1971. New method of collagen extraction for radiocarbon dating. *Nature* 230(5291):241–2.

Lull V, Micó R, Rihuete-Herrada C, Risch R. 2015. When ^{14}C dates fall beyond the limits of uncertainty: an assessment of anomalies in western Mediterranean Bronze Age ^{14}C series. *Radiocarbon* 57(5):1029–40.

Scott EM, Cook GT, Naysmith P, Bryant C, O'Donnell D. 2007. A report on phase 1 of the 5th International Radiocarbon Intercomparison (VIRI). *Radiocarbon* 49(2):409–26.

Scott EM, Cook GT, Naysmith P. 2010a. A report on phase 2 of the Fifth International Radiocarbon Intercomparison (VIRI). *Radiocarbon* 52(3):846–58.

Scott EM, Cook G, Naysmith P. 2010b. The Fifth International Radiocarbon Intercomparison (VIRI): an assessment of laboratory performance in stage 3. *Radiocarbon* 52(3):859–65.

Ward GK, Wilson SR. 1978. Procedures for comparing and combining radiocarbon age determinations: a critique. *Archaeometry* 20(1):19–31.