

14 Quality Control of Groningen ¹⁴C Results from Tel Rehov

Repeatability and intercomparison of Proportional Gas Counting and AMS

Johannes van der Plicht and Hendrik J. Bruins

Abstract

Stratified radiocarbon dates provide a scientific chronological framework independent of cultural assessments. In both Groningen ¹⁴C labs (conventional and Accelerator Mass Spectrometry [AMS]), a total of 64 radiocarbon dates were measured from Tel Rehov, derived from 21 Iron Age loci. This is the largest Iron Age series available at present for any site in the Near East. We present, evaluate, and discuss in this article our methodology in terms of quality assurance, reliability and reproducibility.

Introduction

Radiocarbon dating plays a key role in (pre)historic research, because it provides a scientific yardstick (quite often the only objective) for the measurement of time. This yardstick is independent of cultural deliberations and enables chronological comparisons, for example, of different areas at an excavation site, or between sites and regions. This is essential for proper interpretation of archaeological layers and association with other data (van der Plicht and Bruins 2001).

Our ¹⁴C Iron Age chronology established for Tel Rehov, Israel, unambiguously favors a 'high chronology' (Bruins, van der Plicht, and Mazar 2003a, 2003b). However, our findings are contested by Finkelstein and Piasezky (2003; Chapter 16, this volume). It is of crucial importance to be aware of perceived or real quality problems of dating results. The present Iron Age chronology discussion takes place at the limits of resolution of the ¹⁴C method, as issues need to be resolved with a temporal resolution well within a century. Here even small errors may have important consequences. Mistakes can be made by the ¹⁴C laboratory (methodology and accuracy), and/or in the field (sampling and association).

In this paper we discuss the ¹⁴C dates from Tel Rehov in terms of quality control, as measured by the two Groningen radiocarbon laboratories: conventional by means of Proportional Gas Counting (PGC) and AMS. This is necessary in order to evaluate the coherence and robustness of our radiocarbon series, composed of 64 dates, based on two separate measurement systems (for all these dates, refer to the tables in Mazar *et al.* [Chapter 13, this volume]). Our dates are contested by proponents of the 'low chronology' theory. They refer to other ¹⁴C dates, which are comparatively younger than the Groningen series, as far as Tel Rehov is concerned (Bruins, van der Plicht, and Mazar 2003b, 2004).

The ^{14}C Dating Method

Radiocarbon (^{14}C) is a natural isotope of the element Carbon, which is produced by cosmic radiation and occurs in minute concentrations in living organisms. Since this isotope is radioactive (with a half-life of 5730 years) the ^{14}C concentration gradually declines after the death of the organism. Thus, by measuring the remaining ^{14}C content in (pre)historic samples one can date these samples, or more precisely, calculate the moment of death of the organism.

Although the principle of ^{14}C dating is based on this straightforward model, in practice the theory is complicated by many factors. First, the natural ^{14}C concentration has not been constant through time; the exact value of the half-life to be used for ^{14}C age calculations is not transparent; furthermore, isotopic effects in natural processes change the ^{14}C content for various sample materials, thereby changing the age. Second, measuring the ^{14}C radioactivity in natural concentrations is not simple, and requires specialized laboratories and measuring procedures. Here we summarize briefly the issues relevant for this paper. For more detailed accounts we refer to the specialized literature.

The theoretical complications are solved by a number of conventions (e.g. Mook and Waterbolk 1985). The ^{14}C measurements are reported in a specially defined unit BP. This definition comprises: (a) the ^{14}C radioactivity is measured relative to an international standard (i.e. oxalic acid with a ^{14}C activity corresponding to 1950 AD); (b) the ^{14}C age is calculated using the original half-life value of 5568 years; (c) the ^{14}C age calculation includes correction for isotopic fractionation using the ^{13}C content of the sample (to the standard value of $\delta^{13}\text{C} = -25\text{‰}$).

This convention is valid for all measuring techniques used for ^{14}C dating. Thus, the ^{14}C timescale is *defined*; a ^{14}C chronology (in BP) is different (and varying over time) from a calendar chronology (in BC/AD). The relation between the two timescales is established by calibration, which is the process of translating ^{14}C dates into calendar ages. Calibration curves are established by high precision ^{14}C measurements of wood samples dated absolutely by dendrochronology. The present calibration curve recommended by the ^{14}C community is INTCAL98 (Stuiver *et al.* 1998) to be replaced by INTCAL04 (Reimer *et al.* 2004). The part of the calibration curve relevant for this volume, 1500–500 BCE, is shown in Figure 14.1.

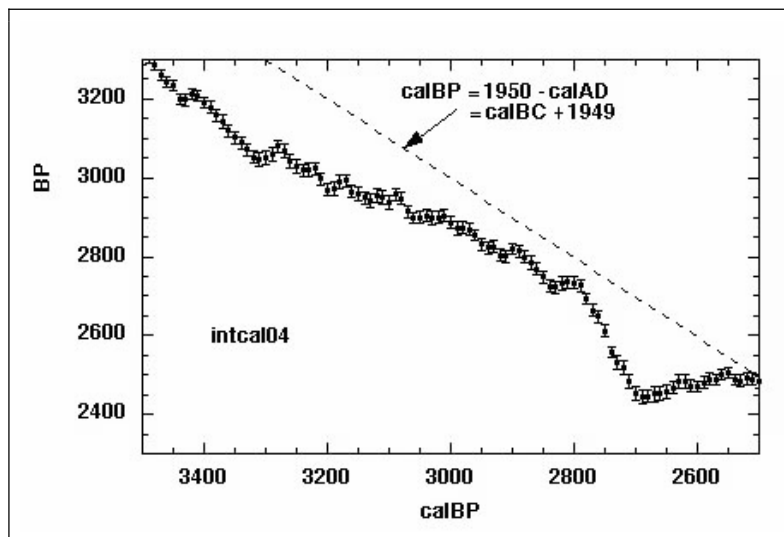


Figure 14.1. Radiocarbon calibration curve (Reimer *et al.* 2004) for the time range 1500–500 BCE.

Calibrated age range distributions for individual or stratified series of ^{14}C dates can be obtained using computer programs like OxCal (Bronk Ramsey 1995).

The measuring process of the ^{14}C content of archaeological samples such as bone, charcoal, wood, and so on, can be viewed from two perspectives: measuring technique and sample preparation. For ^{14}C , three different measuring techniques have been developed—PGC, Liquid Scintillation Spectrometry (LSC), and AMS. The first two, PGC and LSC, are called ‘conventional techniques’ and are based on radiometry; they require ca. 1 gram of carbon (e.g. Kromer and Münnich 1992; Theodorsson 1996). Larger sample amounts may result in a more precise measurement with a smaller standard deviation. The technique of AMS is based on mass spectrometry, for which mg size C is sufficient (e.g. Bayliss, McCormac, and van der Plicht 2004; Tuniz *et al.* 1998). The radiocarbon laboratory in Groningen houses both a conventional (PGC-based) and an AMS laboratory, with both facilities operating independently.

Sample preparation follows similar procedures for all three measuring techniques (Mook and Streurman 1983). The general rules are that contaminants have to be removed (physically and chemically) and that a reliable datable fraction has to be isolated. The latter is then combusted into CO_2 gas, which needs to be purified. For PGC, the $^{14}\text{CO}_2$ is counted in terms of radiometry; for AMS, the CO_2 needs to be transferred into graphite. Stable isotope mass spectrometers are used to measure the $\delta^{13}\text{C}$ content of the same CO_2 gas. Apart from fractionation information, the $\delta^{13}\text{C}$ is a measure for quality of the sample material—as well as the carbon content.

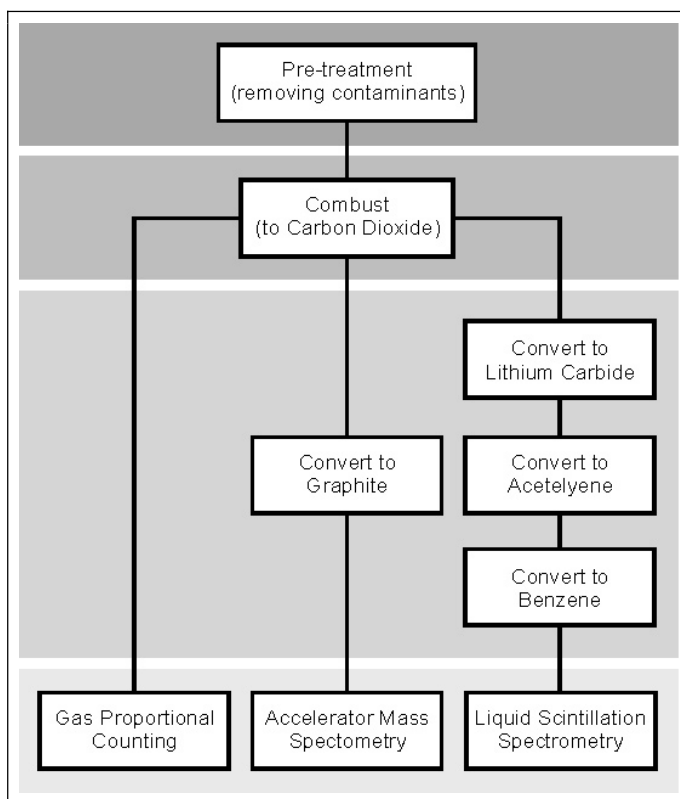


Figure 14.2. Schematic illustration for measuring archaeological samples (Bayliss, McCormac, and van der Plicht 2004; Mook and Streurman 1983): (a) overview of methods for measurements of ^{14}C ; (b) method of pre-treatment for an archaeological charcoal sample.

The pre-treatment for an archaeological charcoal sample is shown as an example in Figure 14.2. Only the ¹⁴C that was part of the organism when it died should be measured. Therefore the first task is to remove any foreign carbon that has entered the sample material since that time. Such contamination comes mainly from the burial environment. We mention the two most common contaminants for charcoal samples: rootlets (usually from higher strata and thus younger) may be present in the collected material; soil carbonate (containing fossil carbon, thus older than the sample). We refer to Mook and Streurman (1983) for more details.

Quality Control Issues

Radiocarbon (¹⁴C) is the most common radiometric dating tool applied in disciplines such as archaeology, geosciences, and environmental research. Stringent quality control is required to build up reliable ¹⁴C chronologies, in particular for the (pre)historical periods in the Levant. Important aspects of quality control involve regular laboratory intercomparisons, multiple analyses of selected samples, sample material selection, archaeological association, and sample size (i.e. conventional versus AMS as measuring technique). These aspects will be discussed below.

General Aspects

True point dates cannot be achieved with ¹⁴C dating, as there will always be a standard deviation. Both equipment resolution and measurement stability, as well as the random nature of radioactive decay (Mook and Waterbolk 1985: 10) ‘causes the results of repeated measurements to spread around a “true” value. The possible discrepancy between a measured value and the “true” value is indicated by the standard deviation (σ)’. Hence one measurement of a sample may result in a date that is very close to the real date, in terms of the mid-point Gaussian value, but may also be older or younger, depending on the value of the standard deviation. Concerning radiocarbon research in Near Eastern historical archaeological periods in which an accuracy and precision is required within a 100 year BP range, multiple measurements of the same sample were suggested (van der Plicht and Bruins 2001). Such multiple measurements will theoretically result in an average date that is both more accurate and precise than can be achieved with single measurements, provided that the ¹⁴C laboratory does not have a systematic bias towards older or younger dates. The issue of multiple measurements is further discussed below.

The quality of the BP date—before calibration—always forms the basis for every radiocarbon age determination. It must be realized that a ¹⁴C date *does* provide a very important universal physical measurement of time, independent of cultural–historical viewpoints and associative reasoning.

Sample selection is a critical component in the ¹⁴C dating process. The layers from which archaeological or geological samples are taken during excavations have not always remained static and may have been affected by different kinds of post-depositional processes. Perturbation by plants, animals or human activities (e.g. digging) may cause migration or contamination of carbon in samples used for ¹⁴C dating.

Another key question is the relationship between the age of the sample and the archaeological or historical question addressed: ‘How is the ¹⁴C event related to the human event to be dated?’ (van Strydonck *et al.* 1999). A well known problem in this respect is the so-called ‘old wood effect’. Wood used (or re-used) to construct a building may have a ¹⁴C date that differs from the human construction event by several centuries, depending on the age of the wood. It must be emphasized that the ¹⁴C date of the wood in such a case is *not* a measurement mistake. Rather the age of the wood sample is older than the age of the archaeological layer or building in which it was found.

Another important matter related to sample selection is the respective choice of ‘conventional dating versus AMS’. There can be a temptation to collect and submit all isolated seeds and tiny flecks of charcoal. The dating of such isolated samples by AMS should be discouraged, if larger samples (seed or charcoal clusters) are present in the same layer. If sufficient material is available, samples can be dated more cheaply and often more accurately by conventional means. The possibility of dating erratic post-depositional influences is considerable when isolated small fragments of charcoal or seeds are used, which are liable to movement by faunal or human digging activity. Such tiny samples have to be derived from a clearly defined context or association to justify dating. Lanting and van der Plicht (1994) presented a detailed discussion about these issues, including examples. It is a ‘myth’ that AMS is better than conventional radiocarbon dating: standard deviations are usually not smaller.

Time-width effects represented by a sample have to be considered. Bulk samples of peat layers, for example, are centimeters thick for conventional ^{14}C analysis. Such a sample comprises many years of sedimentation or growth. Isolated seeds, macrofossils, and grains represent single-year samples and are typical AMS material, due to their small sample size, but the stratigraphic context must be clear, as noted above.

The correct calibration procedure of ^{14}C dates from multi-year or single-year samples needs to be carefully contemplated. Smoothed curves are recommended for multi-year samples, while single-year samples ought to be calibrated with the most detailed calibration curve available (Mook and Waterbolk 1985).

The 975–955 BCE wiggle, so important in the Iron Age (Bruins, van der Plicht, and Mazar 2003a, 2003b; Mazar *et al.* [Chapter 13, this volume]) is based on dendrochronological data through radiocarbon measurements of groups of 10 tree rings (decadal) or 20 tree rings (bidecadal) (Stuiver *et al.* 1998). The radiocarbon measurements of Tel Rehov Stratum VI are based on short-lived samples (seeds, bone) with a time span of a few years or even less (growing season). Past ^{14}C variations in the atmosphere are not captured in such detail by the calibration curve itself, due to the comparatively ‘coarse’ decadal and bidecadal measurements. Therefore, it might be important to make a special primary study of this 975–955 BCE wiggle through ^{14}C measurements of each individual tree ring in the available dendrochronological series for the period 980–940 BCE or even for the entire 10th century BCE. Such detailed measurements have been carried out for older Bronze Age periods (Vogel and van der Plicht 1993).

Intercomparison

Intercomparison is a major part of quality assurance. In an intercomparison, the same samples are dated by different laboratories. This may involve either samples of known age or blind samples. Thus a laboratory can ‘check’ its performance—in particular, the sample (pre)treatment and ^{14}C measurement procedures. Intercomparison is a well recognized issue in the ^{14}C community, and various exercises form a continuing process. The latest large scale intercomparison is FIRI (Fourth International Radiocarbon Intercomparison) in which 84 laboratories participated worldwide. Several publications are generated by this program (Boaretto *et al.* 2003; Scott *et al.* 2004), and the final report is a special publication of the *Radiocarbon* journal (Scott [ed.] 2003).

This FIRI had the following aims:

1. Evaluation of the comparability of routine analysis of both AMS and conventional laboratories.
2. Quantification of the extent of and the sources for any variation.
3. Investigation of the effects of sample size, precision, and pre-treatment on the results (Scott [ed.] 2003).

Groningen Intercomparison and Quality Aspects

We show here the results for the two Groningen laboratories (conventional and AMS) in Table 14.1. The FIRI samples are named A–J. ‘AMS-extra’ were optional samples (not officially part of the FIRI exercise). The consensus value is the resulting value as determined by analysis of the FIRI measurements, except for the wood samples which are dated by dendrochronology.

Table 14.1. FIRI results for the Groningen ¹⁴C laboratories (conventional [lab. code GrN] and AMS [lab. code GrA]). The ¹⁴C ages are reported in BP or as ¹⁴C activities in %. All errors are 1σ. ‘AMS-extra’ were optional samples (not officially part of the FIRI exercise).

	<i>Sample</i>	<i>Conventional</i>	<i>AMS</i>	<i>AMS-extra</i>	<i>Consensus Value</i>
A	Kauri	52400(+1830, -1500)	>50000		0.24%
	cellulose			>50000	
B	Kauri	53140(+2030, -1620)	>50000		0.24%
	cellulose			>50000	
C	Turbidite	18270(70)	17930(60)		18176
D	wood	4510(20)	4485(40)		4508
	cellulose			4470(40)	
E	humic	11850(40)	11800(50)		11780
	1 mg			11860(60)	
	0.2 mg			11650(80)	
	0.06 mg			12050(330)	
F	wood	4490(20)	4525(40)		4508
	cellulose			4550(40)	
G	barley	111.08(0.22)%	110.91(0.31)%		110.7%
H	wood	2190(20)	2240(40)		2232
	cellulose			2230(40)	
I	cellulose	4505(35)	4490(30)		4485
J	barley	110.76(0.22)%	110.72(0.31)%		110.7%

A second class of intercomparison was done by high precision laboratories. This is a ‘subset’ of laboratories, traditionally active in the field of calibration of the radiocarbon timescale (Seattle, Belfast, Heidelberg, Groningen [conventional], Pretoria, and Tucson [conventional]). These laboratories produced the calibration curves, recommended by the ¹⁴C conferences (Reimer *et al.* 2004; Stuiver and Kra [eds.] 1986; Stuiver, Long, and Kra [eds.] 1993; Stuiver and van der Plicht [eds.] 1998). By exchanging dendro-dated wood and standards, the interlaboratory differences yield offsets in the 0–20 ¹⁴C year range (Stuiver *et al.* 1998), and sometimes smaller. The interlaboratory difference between Groningen and Pretoria, for example, appears to be 7.1 ± 6.4 ¹⁴C years, Pretoria being the older of the two. This allowed us to construct a calibration curve with a high temporal resolution (1–4 years) for the 4th–3rd millennia BC (Vogel and van der Plicht 1993).

Thirdly, we mention the development of working standards for the ¹⁴C community. The Groningen laboratory developed the IAEA standards C3, C7 and C8. The latter two have designed activities of 15 and 50%, respectively; C3 is modern (Rozanski *et al.* 1992). The final activities of the material have been measured as an intercomparison between 7 laboratories: Seattle, Belfast, Heidelberg, Groningen (both conventional and AMS), Waikato, and Tucson (conventional). The measured results can be found in LeClercq, van der Plicht, and Gröning (1998).

The Groningen Centre for Isotope Research is in a unique position because it has two radiocarbon laboratories—one conventional and one AMS—which operate independently. Exchange of samples between the two laboratories—as an ‘internal intercomparison’—does take place. Often this happens to test or check the extreme limits of the ¹⁴C method. As an example, we mention the ¹⁴C date of 55,000 BP for a mammoth from Siberia (Arilakh) based on 4 age measurements

(2 conventional, 2 AMS) on 2 types of material (bone marrow and bone collagen). Three of the 4 individual dates would result in an infinite radiocarbon age (Mol *et al.* forthcoming).

More common is the cross-checking of standards and backgrounds. For the conventional laboratory, a few liters of the original standard Oxalic Acid (Ox1) is kept, used, and re-used as the primary standard activity for the conventional laboratory with a radioactivity of 226 ± 1 Bq/kgC (Mook and van der Plicht 1999). The new Oxalic Acid (Ox2) with a ^{14}C activity of 134.06% is used as a working standard for both the conventional and the AMS laboratories.

For AMS, a batch of measurements consists of 57 samples, including typically 5 backgrounds and 8 standards (Aerts, van der Plicht, and Meijer 2001; van der Plicht *et al.* 2000). These standards are Ox2—a mixture of Ox2 combusted by the AMS preparation line following the same route as the unknown samples, and Ox2 from a large volume prepared by the conventional laboratory. This way, the conventional and AMS preparation lines are checked against each other. Similar exchanges between conventional and AMS take place occasionally for background material, which is CO_2 prepared from ^{14}C -free anthracite. We note that the CO_2 gas as prepared by the conventional laboratory is very pure, because this is required for the gas counters (van der Plicht, Streurman, and Schreuder 1992).

In addition, performance of the ^{14}C quality over a long time period is checked using IAEA-C7 and C8. These materials (LeClercq, van der Plicht, and Gröning 1998) are also working standards but treated as unknowns; the Ox2 average ^{14}C measured activity ratio value per batch determines the standard measurement, which is set to 134.06%.

Multiple Analysis

Duplicate measurements of samples may confirm each other, if both results are similar, thus ensuring quality control. But it may also lead to different results. Waterbolk (1990: 148) stated: 'if a sample has been measured twice, be it by the same or by another laboratory, and the results are not congruent, we cannot know which date to reject'. In such cases a third measurement on the same sample should be conducted, in order to determine the correct ^{14}C date.

The standard deviation σ of the radiocarbon date is usually based on the uncertainty in the ^{14}C counts for sample, standard and background. A measure of the reproducibility of the result is another very important factor that influences the standard deviation. Reproducibility is generally estimated by using in-house standards (known age material) and may often be negligible. Nevertheless, this factor is crucial in quality control. Laboratories may be tempted to report a small standard deviation based on radiometry counting only, but neglecting the reproducibility of the date. The latter may also be related to other factors in standard lab routine and performance, including the quality of equipment, and so on.

The calibrated age range is not only defined by counting statistics, but is also 'wobble' dependent. The calibrated age range for a series of dates can be narrowed down by 'wobble matching' (Pearson 1986) or using special statistics (Bronk Ramsey 1998). 'High Precision' measurements are usually defined as those with errors (1σ) < 0.2%. For very large samples containing more than 10 grams of C and measured with Proportional Gas Counting, the smallest feasible limit is 10–15 BP. This lower limit of precision results from the following constraints:

1. The error margins of standards and backgrounds are only slightly smaller.
2. The error margins of the calibration curve are not smaller, as the ^{14}C measurements of the dendro-dated wood are 'high precision' by themselves. The INTCAL98 curve is decadal (i.e. one data-point per 10 tree rings, representing 10 calendar years).

Combined stratified series of short-lived samples and multi-year charcoal are better than only a few isolated short-lived dates (see for examples Bruins and van der Plicht 1995; Bruins *et al.*

concerning Tel Dan [Chapter 19, this volume]). A *paramount requirement* for the establishment of a high-quality radiocarbon chronology of Near Eastern archaeology involves the ¹⁴C data acquisition of as many archaeological layers as possible.

Tel Reḥov: The ¹⁴C Dates

Following an initial series of datings by the ¹⁴C laboratory at Rehovot, Israel (Mazar and Carmi 2001), a series of dates, now reaching 64 (Table 14.2), from 21 loci were measured in Groningen (Bruins, van der Plicht, and Mazar 2003a, 2003b, 2004; Mazar *et al.* [Chapter 13, this volume]). The conventional laboratory in Groningen operates a set of 9 proportional gas counters, with one counter especially suitable for very large samples such as grains (tens of grams). The counter operates for 25 liters of CO₂ and is capable of high precision ¹⁴C dating with errors (1σ) down to 10–15 BP. This enabled the unique high precision dates for sites like Jericho (Bruins and van der Plicht 1995, 2001). Smaller gram-size samples were measured in other proportional counters, and intrinsically small samples were measured by AMS. Thus, both conventional and AMS measurement techniques were used, according to sample size and for quality control reasons. Altogether, the dates for Tel Reḥov constitute the largest group of radiometric measurements from a single Iron Age site in the Levant.

Table 14.2. List of all 64 radiocarbon dates from Tel Reḥov with related stratigraphic and laboratory information. The ¹⁴C dates are measured in Groningen (GrN—conventional [PGC]; GrA—AMS). For a detailed discussion, see Bruins, van der Plicht, and Mazar (2003a, 2004) and particularly Mazar *et al.* (Chapter 13, this volume).

Area, Stratum and Period	Locus	Basket	Charred Organic Material	sample weight ratio	%C organic fraction	Lab. No.	¹⁴ C Date (BP)	δ ¹³ C (‰)
C-1a IV Iron IIA	5498	54702	Cereal Grains	26.9	55.9	GrA-21152	2770 ± 50	–
				same	57.1	GrA-21154	2730 ± 50	22.54
				same	55.6	GrA-21267	2760 ± 35	–
				same	55.5	GrA-22301a	2710 ± 45	22.28
				same	same	GrA-22301b	2775 ± 40	–
				same	55.2	GrA-22330a	2760 ± 50	22.74
				same	same	GrA-22330b	2785 ± 40	–
								22.34
								same
								–
				22.32				
				same				
E-1b ? IV or V ? Iron IIA	2618	46281	Olive stones	36.2	62.0	GrA-17260	2745 ± 40	–
								21.98
B-5 IV or V ? Iron IIA	6229	62430	Total Alkali Alkali Seeds Seeds	20.0	66.3	GrN-28368	2735 ± 30	–
						GrA-24108	2765 ± 45	23.60
						GrA-24109	2770 ± 45	–
						GrA-24111	2780 ± 45	23.60
						GrA-24112	2750 ± 45	–
				–				
				24.30				
				–				
				26.40				

D-2 IV/V/VI Iron IIA	1802	18119	Olive stones	47.1	67.7	GrN-26112	2805 ± 15	- 22.46			
D-3 Iron IB	2862	28493	Olive stones	44.2	58.3	GrA-19033	2835 ± 45	-			
				same	67.1	GrN-26119	2720 ± 30	23.03			
	4815	48105	Olive stones	*	59.7	GrA-16757	2820 ± 50	- 22.51			
				4816	48103	Olive stones	*	57.0	GrA-12889	2870 ± 70	-
						Charcoal	37.0	65.5	GrA-16848	2895 ± 40	25.29
4830	48115	Olive stones	21.6	62.9	GrA-21044	2845 ± 35	-				
			same	62.0	GrA-21056	2825 ± 35	22.05				
			same	63.3	GrA-21183	2820 ± 50	-				
			same	64.4	GrA-22302a	2730 ± 50	23.30				
			same	same	GrA-22302b	2820 ± 40	-				
			same	60.8	GrA-22329a	2810 ± 50	23.35				
			same	same	GrA-22329b	2760 ± 40	-				
									23.00		
						same					
						-					
						22.63					
						same					
D-4 Iron IB	1836	48450	Olive stones	43.6	72.3	GrN-26121	2890 ± 30	-			
				same	58.3	GrA-18825	2870 ± 50	22.95			
D-6 Iron IA	1845	48556	Seeds	21.6	61.4	GrA-21046	2905 ± 35	-			
				same	61.8	GrA-21057	2945 ± 35	22.49			
	2836	28352	Olive stones	same	60.8	GrA-21184	2920 ± 50	-			
								23.10			
						-					
						24.12					
D-6 Iron IA	2836	28352	Olive stones	49.4	68.3	GrN-26118	2920 ± 30	-			
				same	54.8	GrA-18826	2950 ± 50	22.28			
D-6 Iron IA	2874	28701	Olive stones	38.0	58.9	GrA-19034	2935 ± 45	-			
				same	72.9	GrN-26120	2880 ± 30	22.14			
						-					
						22.36					

* = pre-treatment conventional, measured by AMS.

The list of all 64 ¹⁴C dates are shown in stratigraphic order in Table 14.2 (see also Mazar *et al.* [Chapter 13, this volume], for more archaeological data and the calibration of the dates). The Tel Rehov dates measured in the Rehovot ¹⁴C laboratory (RT) (Gilboa and Sharon 2004; Mazar and Carmi 2001) are shown in the tables in Mazar *et al.* (Chapter 13, this volume). It appears from these tables that there is a systematic difference between RT and GrN/GrA, as RT is consistently younger. This is a significant observation, because the low/high chronology debate is partly caused by this interlaboratory deviation. Examples are discussed below.

A sample of charred olive pits from the oldest Iron I Stratum in Area D (D-6, Iron IA) was dated in Rehovot to 2685 ± 40 BP (RT-3119); this result is much too young by all standards as already noted by Mazar and Carmi (2001). The same stratum also yielded 4 Groningen dates

(2 conventional, 2 AMS), which are older by more than 200 BP (GrA-18826, 19034; GrN-26118, 26120).

Seeds from phase D-4b were dated at Rehovot to 2800 ± 40 BP (RT-3121), while the same seeds were dated in triplicate by AMS in Groningen yielding consistently older (ca. 100 BP) results (GrA-21046, 21057, 21184).

Another sample of olive stones from phase D-3 was dated at Rehovot to 2670 ± 40 BP (RT-3120), which is also considered much too young (Mazar and Carmi 2001).

We note that most Tel Rehov samples measured in Groningen originate from a clear archaeological context, that is, there is no association problem. All samples (seeds, olive pits) represent truly single year material. The entire series forms an excellent set of material from the ^{14}C point of view. Table 14.2 shows the Groningen dates (GrN and GrA) for the Iron Age IIA strata, in relation to the stratigraphy. The archaeological interpretations are discussed elsewhere (Bruins *et al.* 2004; Bruins *et al.* [Chapter 19, this volume]; Mazar *et al.* [Chapter 13, this volume]).

The strong point of the Groningen set of dates (Table 14.3) is the quality control aspect; two independent laboratories, one conventional including a high precision counter for large quantities of single year sample material, and one AMS. The AMS cannot achieve high precision but triplicate analyses were performed (both pre-treatment and measurement were performed in triplicate!). This enables the calculation of a weighted average of the dates in order to increase the precision and possibly also the accuracy, resulting in a measurement that is closer to the 'true age'.

Table 14.3. Specific tests for the ^{14}C methodology of additional Groningen Radiocarbon dates (GrN—conventional; GrA—AMS) for Tel Rehov.

Phase and Stratum	Locus	Basket	Charred Organic Material	Lab. No.	^{14}C Date (BP)	$\delta^{13}\text{C}$ (‰) Fraction
C-1a IV	5498	54702	Charred seeds	GrA-22301	2710(45)	-22.34
				GrA-22301	2775(40)	same
				GrA-22330	2760(50)	-22.32
				GrA-22330	2785(40)	same
D-3	4830	48115	Olive pits	GrA-22302	2730(50)	-23.00
				GrA-22302	2820(40)	same
				GrA-22329	2810(50)	-22.63
				GrA-22329	2760(40)	same
C V	2441	24579	Cereal grains	GrN-27363	2745(15)	-22.66 coarse
				GrN-27385	2770(15)	-22.31 middle
				GrN-27386	2760(15)	-22.37 fine
C-1b V	2444	24647	Cereal grains	GrN-27364	2765(15)	-22.15 coarse
				GrN-27413	2865(20)	-22.04 fine
C-2 VI	4426	44166	Bone	GrA-21417	2840(45)	-19.64 collagen
	6229	62430	Charred seeds	GrA-24108	2765(45)	-23.60 alkali
				GrA-24109	2770(45)	-23.60 alkali
				GrA-24111	2780(45)	-24.30 seeds
				GrA-24112	2750(45)	-26.40 seeds
			GrN-28368	2735(30)	-23.60 total	
B-5 V b	4218	42236	Charcoal	GrA-21042	2765(35)	-23.80
				GrA-21053	2750(35)	-25.62
				GrA-21180	2690(50)	-26.36

Moreover, this (costly) exercise is also a quality check: practically all individual dates of the same sample material (basket) are within 1σ of each other. Such consistent results are indicative of high quality with robust repeatability (Scott [ed.] 2003).

In addition, the two Groningen techniques (conventional and AMS) enable internal intercomparison. Occasionally, samples were measured by both techniques, yielding consistent results. This is in accordance with other internal quality procedures which are performed routinely, and is a strong indication for quality and reliability (Scott [ed.] 2003).

Table 14.2 shows that many samples are dated by multiple analyses. The duplicates overlap very well, mostly within 1σ , which justifies the calculation of weighted averages. This applies to both AMS (typical single date errors 35–50 BP) and conventional (mostly high precision, errors ≈ 15 BP). Therefore, the Groningen ¹⁴C results from Tel Rehov constitute a high quality dataset from all physical/chemical perspectives.

Table 14.3 highlights additional dates, measured in a later stage as a test exercise for the ¹⁴C methodology. The first eight measurements (baskets 54702 and 48115) constitute a ‘double duplication’ test: charred seeds and olive pits are treated in duplicate (GrA-22301/22330 and GrA-22302/22329), and these four samples were measured again at a later time (AMS measurement duplicate, same graphite targets). All measured ¹⁴C dates show excellent reproducibility. Note that one set (basket 54702) had been dated earlier in triplicate as well (Table 14.2), yielding the same ¹⁴C dates within error.

Homogeneity tests were performed for samples from baskets 24579 and 24647. This concerns large samples, measured by the conventional laboratory. The sample material was divided into parts, according to size fractions, by sieving: coarse ($>850\ \mu\text{m}$), middle ($<850\ \mu\text{m}$, $>180\ \mu\text{m}$) and fine ($<180\ \mu\text{m}$). The coarse fraction is probably the most reliable in terms of ¹⁴C analysis; the recognizable seeds are obviously present in this fraction. The fine fraction may contain charred soil organic matter, and if so, it should be older. Indeed, for basket 24647, GrN-27413 is older than GrN-27364. It was suggested by Mazar *et al.* (Chapter 13, this volume) that the fine material indeed is probably derived from charred wooden beams.

However, the other samples that were split in terms of size fractions do not show an age difference. For basket 24579 the three fractions yield the same ¹⁴C date. Note that also basket 24408b (Table 14.2) consists of such a measurement—GrN-27412 corresponds to the fine fraction, and GrN-27362 to the coarse fraction of the same sample material. Also here the ¹⁴C dates are the same within the σ error.

Concerning basket 44166, besides the charred seeds, a bone sample was available for dating. Collagen is extracted from the bone material as the datable fraction (Mook and Streurman 1983). Both types of organic material yielded the same ¹⁴C date, within the error.

Finally, a pre-treatment test was performed. Seeds from basket 62430 were dated both in terms of the alkali fraction and the residue (the ‘normal’ fraction). The chemical part of this experiment was done by the conventional laboratory. However, the separate alkali and residue fractions appeared too small for a precise conventional radiometric measurement. Hence they were sampled and transferred to the AMS laboratory for further analysis. At the end, both fractions were put together again and measured by the conventional laboratory (total fraction, GrN-28368). The AMS measurements were done in duplicate to test reproducibility. The overall total of five ¹⁴C measurements is in excellent agreement in physical terms.

Problem Analysis—Is there a Way Out ?

We believe that our ¹⁴C methodology, as applied to Tel Rehov, is robust and coherent, representing the best that ¹⁴C has to offer by using stringent quality control. Nevertheless, some problems

remain to be solved. For example, there appears to be a significant offset between the ^{14}C dates from Groningen and the earlier dates obtained by Rehovot. Below is a list of possible causes for 'wrong' ^{14}C dates (e.g. van Strydonck *et al.* 1999):

1. A wrong archaeological association in the field.
2. Mislabeling of samples (in the field, storehouse or ^{14}C laboratory).
3. Contamination of the sample (in the field or in the laboratory).
4. The sample is not homogeneous or otherwise problematic.
5. The ^{14}C measurement is wrong (e.g. standards, background problems).
6. The sample pre-treatment is erroneous or not adequate.
7. The fractionation ($\delta^{13}\text{C}$ value) correction is not correct.
8. A combination of the above.

We note that all of the above mishaps do occasionally happen, as the volume of samples going through a ^{14}C laboratory can be very large (e.g. in Groningen, around 4000 samples annually). As an illustrative example, we can calculate (Mook and van der Plicht 1999) that an error of around 100 BP for a Levantine Iron Age sample is introduced by an error of 6‰ in $\delta^{13}\text{C}$, or by contamination with 1% fossil (^{14}C free) or 3% modern material.

The $\delta^{13}\text{C}$ value serves as a quality parameter for the sample material. It can be seen from the measurements listed in Tables 14.2 and 14.3 that for Tel Rehov, these values are usually in the -22 to -23 range for seeds, and around -24 for charcoal (see also Bruins *et al.* concerning Tel Dan [Chapter 19, this volume]). The one bone sample (GrA-21417) clearly has a different $\delta^{13}\text{C} = -19.64\text{‰}$, which is a normal value for bone material.

The second quality parameter used in the ^{14}C laboratory is the organic content or the carbon content of the sample. In Table 14.2, two numbers are listed. The first column states the weight ratio (in %) of the original sample after/before the pre-treatment procedure. The carbon content of the organic fraction (%C), however, is the best sample material quality parameter. For seeds and charcoal, this value should be around 60%. From this perspective, all samples dated qualify as good datable material. Note that some samples were routed in an unusual way: for example, GrA-16757 and GrA-12889 (*) were pretreated by the conventional laboratory, but combusted and measured by the AMS laboratory.

Besides the deviating dates from Rehovot, there is one Groningen date which is an outlier: GrN-26119 from Stratum D-3 is too young, also by about a century (Table 14.2). This particular sample was submitted as a conventional sample, but appeared too small and hence was rerouted to the AMS laboratory (GrA-19033). Nevertheless, the sample could be measured conventionally by adding fossil (^{14}C free) CO_2 (Mook and Streurman 1983). We suspect the source of error is caused by this procedure. If so, this qualifies as a 'systematic error'.

The errors quoted for ^{14}C measurements only include measurement errors, which are determined from statistics (i.e. the number of ^{14}C counts), and scatter of standards and/or backgrounds. Other errors are difficult to quantify because they are not known; they can be assessed by sample quality parameters like the $\delta^{13}\text{C}$ value and organic carbon content, or tests concerning homogeneity and pre-treatment (see examples in Table 14.3). Multiple analysis is a good way to test reproducibility and reliability; intercomparison between different laboratories is a way to resolve possible discrepancies. This is one of the goals of intercomparison exercises like FIRI. As for Tel Rehov, an intercomparison between Rehovot/Groningen/Tucson is underway for current measurements. This intercomparison will not necessarily be able to resolve the reason for the rather young dates of the Rehovot lab, measured during the 1990s, in comparison to Groningen dates from the same material or strata from Tel Rehov.

Conclusions

In general, it is obvious that quality control in ^{14}C dating is essential in order to obtain meaningful results in the historical archaeological periods of the Near East (van der Plicht and Bruins 2001; Scott [ed.] 2003). Both archaeological and scientific elements are involved. The Groningen dates of Tel Rehov come from a detailed stratigraphic series of reliable sequenced contexts, each providing short-lived samples. These are perfect materials for dating from the radiocarbon point of perspective (van Strydonck *et al.* 1999). Internal quality assessment and internal intercomparison were performed between the two independent Groningen laboratories (conventional: PGC and AMS). The repeatability of the various dates is generally excellent, with only very few exceptions.

Reproducibility of a radiocarbon measurement on the same sample is both a check and confirmation of accuracy and reliability. It is, therefore, important that key samples are measured to the highest possible precision, that is, they should be short-lived, large in size and preferably subject to duplicate or triplicate analysis. In some cases, the same samples or loci were dated earlier by a third laboratory, Rehovot. Our research resulted, therefore, in an unplanned intercomparison with this laboratory. It appears that for Tel Rehov, the results from the Rehovot lab are systematically younger than those from both Groningen laboratories.

Since the consistent Groningen dates are based on both PGC and AMS measurements, we are inclined to consider them as reliable, while the Rehovot dates from Tel Rehov are too young by a century or more. The reason for the apparent systematic offset between the Radiocarbon laboratories of Rehovot and Groningen remains to be solved. An independent intercomparison for the current lab operation in 2004 is underway. Ultimately, in order for the Rehovot lab to check the comparatively low dates mentioned above, it would be necessary for the lab to date this material again in its current operational setup and working procedures in order to evaluate the possible problems in former measurements.

References

- Aerts, A.T., J. van der Plicht, and H.A.J. Meijer (2001) Automatic AMS Sample Combustion and CO_2 Collection. *Radiocarbon* 43: 293-98.
- Bayliss, A., F.G. McCormac, and J. van der Plicht (2004) An Illustrated Guide to Measuring Radiocarbon from Archaeological Samples. *Physics Education* 39: 137-44.
- Boaretto, E., *et al.* (2003) How Reliable are Radiocarbon Dates? A Report on the Fourth International Radiocarbon Intercomparison (FIRI) (1998–2001). *Antiquity* 295: 146-54.
- Bronk Ramsey, C. (1995) Radiocarbon Calibration and Analysis of Stratigraphy: The OxCal Program. *Radiocarbon* 37: 425-30.
- (1998) Probability and Dating. *Radiocarbon* 40: 461-74.
- Bruins, H.J., and J. van der Plicht (1995) Tell Es-Sultan (Jericho): Radiocarbon Results of Short-lived Cereal and Multi-year Charcoal Samples from the End of the Middle Bronze Age. *Radiocarbon* 37: 213-20.
- (2001) Radiocarbon Challenges Archaeo-historical Time Frameworks in the Near East: The Early Bronze Age of Jericho in Relation to Egypt. *Radiocarbon* 43: 1321-32.
- Bruins, H.J., J. van der Plicht, and A. Mazar (2003a) ^{14}C Dates from Tel Rehov: Iron-Age Chronology, Pharaohs, and Hebrew Kings. *Science* 300: 315-18.
- (2003b) Response to Comment on ^{14}C Dates from Tel Rehov: Iron-Age Chronology, Pharaohs, and Hebrew Kings'. *Science* 302: 568c-d.
- (2004) The End of the 2nd Millennium BCE and the Transition from Iron I to Iron IIA: Radiocarbon Dates of Tel Rehov, Israel (Vienna: SCIEM).
- Finkelstein, I., and E. Piasezky (2003) Comment on ^{14}C Dates from Tel Rehov: Iron-Age Chronology, Pharaohs, and Hebrew Kings'. *Science* 302: 568b.

- Gilboa, A., and I. Sharon (2004) An Archaeological Contribution to the Early Iron Age Debate: Alternative Chronologies for Phoenicia and their effects on the Levant, Cyprus, and Greece. *BASOR* 322: 7-80.
- Kromer, B., and K.O. Münnich (1992) CO₂ Gas Proportional Counting in Radiocarbon Dating—Review and Perspective. In *Radiocarbon after Four Decades: An Interdisciplinary Perspective*, edited by R.E. Taylor, A. Long, and R.S. Kra (New York: Springer Verlag): 184-97.
- Lanting, J.N., and J. van der Plicht (1994) ¹⁴C-AMS: Pros and Cons for Archaeology. *Palaeohistoria* 35/36: 1-12.
- LeClercq, M., J. van der Plicht, and M. Gröning (1998) New ¹⁴C Reference Materials with Activities of 15 and 50 pMC. *Radiocarbon* 40: 295-97.
- Mazar, A., and I. Carmi (2001) Radiocarbon Dates from Iron Age Strata at tel Beth Shean and Tel Rehov. *Radiocarbon* 43: 1333-42.
- Mol, D., *et al.* (forthcoming) Results of the Cerpolex/Mammothus Expeditions on the Taimir Peninsula, Arctic Siberia, Russian Federation. *Quaternary International* 142-43: 186-202.
- Mook, W.G., and H.J. Streurman (1983) Physical and Chemical Aspects of Radiocarbon Dating. In *Proceedings of the First International Symposium ¹⁴C and Archaeology, Groningen*, edited by W.G. Mook and H.T. Waterbolk (PACT 8; Strasbourg: Council of Europe): 31-55
- Mook, W.G., and J. van der Plicht (1999) Reporting ¹⁴C Activities and Concentrations. *Radiocarbon* 41: 227-39.
- Mook, W.G., and H.T. Waterbolk (1985) *Handbook for Archaeologists. III. Radiocarbon Dating* (Strasbourg: European Science Foundation).
- Pearson, G.W. (1986) Precise Calendrical Dating of known Growth-period Samples using a 'Curve Fitting' Technique. *Radiocarbon* 28: 292-99.
- Reimer, P.J., *et al.* (2004) INTCAL04 Terrestrial Radiocarbon Age Calibration, 26-0 ka BP. *Radiocarbon* 46: 1029-58.
- Rozanski, K., *et al.* (1992) The IAEA ¹⁴C Intercomparison Exercise. *Radiocarbon* 34: 506-19.
- Scott, E.M. (ed.) (2003) The Third International Radiocarbon Intercomparison (TIRI) and the Fourth International Radiocarbon Intercomparison (FIRI). *Radiocarbon* 45: 135-408.
- Scott, E.M., *et al.* (2004) Precision and Accuracy in Applied ¹⁴C Dating: Some Findings from the 4th International Radiocarbon Comparison. *Journal of Archaeological Science* 31: 1209-13.
- Strydonck, M. van, *et al.* (1999). What's in a ¹⁴C Date. *Proceedings of the Third International Symposium ¹⁴C and Archaeology, Lyon, France, 6-10 April 1998* (Paris: Société Préhistorique Française): 433-448.
- Stuiver, M., and R.S. Kra (eds.) (1986) Calibration Issue. *Radiocarbon* 28(2B): 805-1030.
- Stuiver, M., A. Long, and R.S. Kra (eds.) (1993) Calibration Issue. *Radiocarbon* 35(1).
- Stuiver, M., and J. van der Plicht (eds.) (1998) INTCAL98, Calibration Issue. *Radiocarbon* 40(3).
- Stuiver, M., *et al.* (1998) INTCAL98 Radiocarbon Age Calibration, 24,000-0 cal BP. *Radiocarbon* 40: 1041-84.
- Theodorsson, P. (1996) *Measurement of Weak Radioactivity* (Singapore: World Scientific Publishing).
- Tuniz, C., *et al.* (1998). *Accelerator Mass Spectrometry* (Boston/London/New York/Washington, DC: CRC Press).
- Van der Plicht, J., and H.J. Bruins (2001) Radiocarbon Dating in Near-Eastern Contexts: Confusion and Quality Control. *Radiocarbon* 43: 1155-66.
- Van der Plicht, J., H.J. Streurman, and G.R. Schreuder (1992) A New Data Acquisition System for the Groningen Counters. *Radiocarbon* 34: 500-505.
- Van der Plicht, J., *et al.* (2000) The Groningen AMS Facility: Status Report. *Nuclear Instruments and Methods B* 172: 58-65.
- Vogel, J.C., and J. van der Plicht (1993) Calibration Curve for Short-lived Samples, 1900-3900 BC. *Radiocarbon* 35: 87-91.
- Waterbolk, H.T. (1990) Quality Differences between Radiocarbon Laboratories Illustrated on Material from SW Asia and Egypt. *Proceedings of the Second International Symposium ¹⁴C and Archaeology, Groningen*, edited by W.G. Mook and H.T. Waterbolk (PACT: Journal of the European Study Group on Physical, Chemical and Mathematical Techniques Applied to Archaeology 29; Strasbourg: Council of Europe): 141-58.