

Bayesian Statistics in Radiocarbon Calibration[‡]

Daniel Steel

Department of History and Philosophy of Science

University of Pittsburgh

Word Count: 4,981

[‡]I would like to thank Clark Glymour, Deborah Mayo, and John Earman for helpful comments on earlier drafts of this essay.

ABSTRACT

Critics of Bayesianism often assert that scientists are not Bayesians. The widespread use of Bayesian statistics in the field of radiocarbon calibration is discussed in relation to this charge. This case study illustrates the willingness of scientists to use Bayesian statistics when the approach offers some advantage, while continuing to use orthodox methods in other contexts. Such an eclectic and pragmatic use of statistical methods casts doubt on inferences from the use of a method to a strong commitment to principles underlying the method.

1. Introduction. According to Deborah Mayo:

Scientists do not proceed to appraise claims by explicit application of Bayesian methods. They do not, for example, report results by reporting their posterior probability assignments to one hypothesis compared with others—even dyed-in-the-wool Bayesians apparently grant this. (1996, 89)

The charge that real scientists are not Bayesians does indeed have a good deal of rhetorical force. In this essay, I discuss the claim in relation to a field, radiocarbon calibration, in which Bayesian methods are not only explicitly advocated by some researchers but are also employed in the standard statistical software package, called CALIB.

I argue that, given the most charitable reading, the “scientists are not Bayesians” argument aims to establish that scientists generally accept norms concerning proper scientific inference that contradict the principles underlying Bayesian statistics. However, the use of Bayesian statistics in radiocarbon calibration calls this argument into question, and not only in virtue of being a counter instance to the summary generalization upon which the argument is premised. More importantly, the present example reveals the willingness of scientists to use Bayesian methods when they perceive some advantage from doing so, while paying little if any heed to the fundamental conflicts of principle between Bayesian and frequentist statistics. This clearly suggests that the use of a method need not involve any strong commitment to, or even awareness of, the principles typically invoked to justify it, which, I argue, undermines the central inference of the “scientists are not Bayesians” argument.

2. Scientists are not Bayesians. The claim that scientists are not Bayesians could be interpreted in more than one way. One strategy would be to argue that scientists, and people in general, reason in ways that systematically violate the canons of Bayesian inference (cf. Giere 1988, 153). Thus interpreted, however, the charge has little force as a criticism, since Bayesianism is a *normative* theory. It would be absurd, for example, object to deductive logic on the grounds that people are systematically inclined to commit logical fallacies (cf. Howson and Urbach 1993, 422-3). In short, that people do not always what they should is no objection to a proposed normative principle.

A better way to put the “scientists are not Bayesians” argument would be to argue that the dominant use of non-Bayesian statistical methods is an indication that scientists generally accept norms of inductive inference sharply at odds with Bayesian principles. Given this formulation, the point of the argument is not just that scientists fail to abide by Bayesian principles but that scientists do not think that they *should* follow them. It is one thing to break a commandment, it is another to reject the legitimacy of the commandment itself. The “scientists are not Bayesians” argument, charitably interpreted, aims to show that it is in this latter sense that scientists fail to be Bayesian. Moreover, some Bayesians wish to deny this conclusion:

It is not prejudicial to the conjecture that *what we ourselves take to be correct inductive reasoning* is Bayesian in character that there should be observable and sometimes systematic deviations from Bayesian precepts. (Howson and Urbach 1993, 423; italics in original)

The denial that overt violations of “Bayesian precepts” demonstrate a lack of agreement with these precepts suggests that some Bayesians think that it would reflect badly on their view if scientists in general rejected its underlying principles.

Using the expression “frequentist statistics” to refer the non-Bayesian statistical methods commonly used in science, the “scientists are not Bayesians” argument has the following structure:

(P1) Frequentist statistics are the most commonly used statistical methods in science.

∴ (C1) Scientists in general accept the principles underlying frequentist statistics.

(P2) The principles underlying frequentist statistics are inconsistent with the principles underlying Bayesian statistics.

∴ (C2) Scientists in general reject the principles underlying Bayesian statistics.

My primary objection to this argument will concern the inference from (P1) to (C1).

It will be useful to note briefly the conflicts of principle between Bayesian and frequentist statistics that are relevant to the present case. Bayesianism is founded on the claim that (ideally) rational people have degrees of belief that are consistent with the probability calculus, and that scientific inference is a matter of changing these degrees of belief in accordance with Bayes’ theorem as new information is received. Suppose that $\{H_1, H_2, \dots\}$ is a set of mutually exclusive and collectively exhaustive hypotheses, and let H be some member of this set. Then Bayes’ theorem can be stated as follows:

$$P(H | E) = \frac{P(H) \times P(E | H)}{\sum P(H_i) P(E | H_i)}, \text{ where } i = 1, 2, \dots$$

$P(H | E)$ is called the *posterior probability*, $P(H)$ the *prior probability*, and $P(E | H)$ the *likelihood*. Typically, it is assumed that after the data E has been learned, the posterior probability becomes the new prior probability of H . Hence, the probabilities represent the beliefs of (some ideally rational) person at a given time: as new information is acquired, yesterday's posterior probability becomes today's prior.

Some points about terminology will be helpful. The *prior probability distribution* is a function that tells us the value of $P(H_i)$, for any i . Likewise, the *posterior probability distribution* is a function specifying the value of $P(H_i | E)$, for any i . The *likelihood function* tells us the value of $P(E | H_i)$, for any i . Notice that given the prior probability distribution and likelihood function, the posterior distribution can be computed by Bayes' theorem. One type of prior probability distribution that will be frequently mentioned below is the *uniform prior probability distribution*. The prior probability distribution is said to be uniform just in case there is some constant k such that $P(H_i) = k$, for all i .

The principles underlying frequentist methods conflict with the Bayesian approach in a variety of ways. The most obvious difference is that probabilities enter frequentist statistics not as degrees of belief in hypotheses, but as estimates of the relative frequencies with which a method produces certain types of results (cf. Mayo 1996, 9-10, 352). For example, estimates in classical statistics are often expressed via *confidence intervals* and associated *confidence coefficients*. In the case of a radiocarbon date, the confidence interval might be 5000 ± 100 years before present and the confidence

coefficient 95%. The confidence coefficient is not, as might be thought, the probability that the age of the item is between 5100 and 4900 years before present.¹ Rather, the confidence coefficient tells us something about the method by which the estimate was generated (cf. Lindgren 1993, 282-3). That is, in a large number of repetitions, the method would very likely produce an estimate within 100 years of the correct date with a relative frequency of approximately .95.

Suppose our method generates as a point estimate m associated with some range of error, such as $m - \partial \leq \theta \leq m + \partial$, where θ is the parameter being estimated. Frequentist statistics demands that such a method be *unbiased*, *efficient*, and *consistent*. An estimation method is unbiased just in case the expected value of its estimate equals θ , the parameter of interest. Efficiency is a measure of how probable it is that the estimate will be within a given range of its expected value: the more probable that the estimate will be close to its expected value, the greater the efficiency. The appeal of unbiased and efficient estimators is easy to see, since an estimator that is unbiased and efficient is highly likely to produce an estimate close to θ . Consistency means that, for each possible value r of θ , the estimate produced by our method converges to r in the limit with probability 1 if $\theta = r$.

Bayesian estimation methods may easily fail to be unbiased in the sense defined above. For example, suppose we have some information makes it reasonable to suppose that some values of θ are more probable than others. Then according to the Bayesian approach, our prior probability distribution must reflect this information. However, given such a prior distribution, Bayes' theorem may be a biased estimator in the classical sense

¹ See Howson and Urbach (1993, 239-41) for an explanation of why such an interpretation is not admissible within a frequentist context.

(e.g., if the mean of the prior distribution differs from θ). In general, Bayes' theorem can be assured of being an unbiased estimator only given a uniform prior probability distribution.

3. Some Basics of Radiocarbon Calibration. Radiocarbon dating is based on the decay of the radioactive isotope carbon 14 (^{14}C) that occurs with a half-life whose currently accepted value is 5,568 years (Stuiver and Polach 1977). ^{14}C is continually created in the upper atmosphere as a result of collisions among neutrons and nitrogen atoms, and some of the ^{14}C atoms join with oxygen to form $^{14}\text{CO}_2$ molecules (Campbell and Loy 1996, 26-7). Such $^{14}\text{CO}_2$ molecules are absorbed by plants through the process of photosynthesis, which breaks down the molecule into ^{14}C , which is used in the construction of cells, and O_2 , which is released into the atmosphere. Thus, the plant comes to contain a ratio of ^{14}C to the stable carbon 12 (^{12}C) that corresponds to the ratio of $^{14}\text{CO}_2$ to $^{12}\text{CO}_2$ in the atmosphere, and this ratio is then passed along through the food chain. Once the plant or animal dies, however, it no longer absorbs additional ^{14}C , and the proportion of ^{14}C to ^{12}C gradually decreases over time as ^{14}C atoms decay. Given estimates of the half-life of ^{14}C and the ratio of $^{14}\text{CO}_2$ to $^{12}\text{CO}_2$ in the atmosphere, it is then possible to make an estimate at the date at which the plant or animal died.

However, this simple picture is made enormously more complex by the fact that the proportion of atmospheric $^{14}\text{CO}_2$ to $^{12}\text{CO}_2$ is known to have fluctuated throughout history and to be sensitive to certain local conditions, such as volcanic eruptions. Consequently, within the last thirty years a great deal of work in the field of radiocarbon dating has been consecrated to what is called "calibration," by which is meant the

following process. First, estimate the age of the sample on the assumption that the current atmospheric $^{14}\text{CO}_2$ to $^{12}\text{CO}_2$ ratio has remained constant throughout the earth's history, and call this estimate the "radiocarbon age" of the sample (Stuiver and Pearson 1992, 19). Next, adjust this estimate by taking into account the historical fluctuations of atmospheric $^{14}\text{CO}_2$ to $^{12}\text{CO}_2$, and call the adjusted estimate the "calendar age" of the sample. "Calibration" is the name for the process of deriving the calendar age from the radiocarbon age.

The strategy used to develop a reliable means of calibration has been to find organic materials, which can be radiocarbon dated, and which can be independently dated by some other method. By correlating radiocarbon to independently obtained calendar dates, researchers have created a "calibration curve" that allows radiocarbon dates to be properly adjusted. Initially, the calibration curve was constructed from timber samples that could be reliably dated by dendrochronology (i.e., tree-ring dating) (cf. Stuiver and Pearson 1993). More recently, the time range of the calibration curve has been greatly extended through the use of corals, which can be dated on the basis of uranium/thorium decay as well as ^{14}C dating (Stuiver et al. 1998). At present, calibration curve stretches from about zero to 24,000 years before present² (ibid. 1041).

In the simplest case, radiocarbon calibration involves transferring a single radiocarbon date into an estimate of a calendar date, but calibration often involves more complex problems (Buck et al. 1991, 812; Pazdur and Michczynska 1989, 824). For example, it is often desirable to combine several radiocarbon dates into one calendar date estimate. Moreover, prior knowledge sometimes makes it possible to place absolute limits on the range of possible ages of the sample or samples being dated. Another

situation involves what is known in archeology as “phasing.” Archeological sites often exhibit distinct strata corresponding to a series of occupations of the same location, and the term “phasing” refers to the process of reconstructing these stages. The term “phase” is used to refer to such occupational stages within a site. Naturally, estimating the beginning and ending dates of the phases of sites is a common concern of archeologists. Moreover, archeological excavation typically provides important information on this score, since we may reasonably assume that strata further down in the excavation are older than higher strata.

3. The Calibration Curve and CALIB. The most important feature of the calibration curve, for our purposes, is that it is very “wiggly,” which has the result that one radiocarbon date may correspond to several calendar dates. Thus, a probability distribution associated with a calibrated radiocarbon date is usually multi-peaked. Moreover, the process of calibration begins with a radiocarbon date, which is itself an estimate associated with a probability distribution. In a review of developments in radiocarbon calibration, Bowman and Leese state:

A consensus is emerging ... that calibrated dates can be faithfully represented only by probability distributions that fully take account of both the error term on the radiocarbon result and the effect of the wiggles in the curve; the wiggles indicate that any one radiocarbon date can correspond to more than one calendar age. (1995, 104)

The first statistical program designed for radiocarbon calibration that I know of utilized a common type of orthodox statistical method known as a maximum likelihood

² Where “present” is fixed at 1950 by convention.

estimator (Orton 1983).³ However, this program failed to become widely used, primarily because of certain computational difficulties involved in its implementation that apparently arose from the irregular shape of the calibration curve (Naylor and Smith 1988, 591; Buck et al. 1991, 812-3). The first version of CALIB was published in 1986 in *Radiocarbon*, the main journal in the field (Stuiver and Reimer 1986). Since then CALIB has become, “The most well-known and commonly used computer program ...for the calibration of radiocarbon determinations” (Buck et al. 1996, 215). A special issue of the *Radiocarbon* in 1993 dedicated to calibration came equipped with a diskette bearing CALIB. A 1998 issue of *Radiocarbon* (volume 40, number 3), whose centerpiece is the presentation of an updated calibration curve, provides the address to a web-site from which the latest version of CALIB can be downloaded.⁴

Let us consider how CALIB deals with the simplest type of case, in which we wish to infer a calendar date from a single radiocarbon estimate (cf. Buck et al. 1996, 212-5). Here the data is a particular radiocarbon estimate (say, 5000 years BP). The hypotheses assert different possible calendar ages for the item being dated. Hence, we want to compute a posterior probability distribution for these hypotheses given the radiocarbon estimate.

Recall that we can compute a posterior probability distribution via Bayes’ theorem given a prior probability distribution and a likelihood function. CALIB assumes that the prior distribution is uniform over the calendar dates. The likelihood function is computed in the following way. Suppose that the calendar date of the item in question is t , and let $\mu(t)$ represent the corresponding radiocarbon date given by the calibration

³ Maximum likelihood estimators produce as their estimate the hypothesis that maximizes the probability of the data.

curve. Then assuming that the calibration curve is accurate, the correct radiocarbon date is $\mu(t)$. But of course, the process of generating radiocarbon estimates involves a degree of random error, so the particular estimate produced on a given occasion is likely to differ from the correct value. Let X be a random variable that ranges over possible values of radiocarbon estimates. It is then assumed that the distribution of X conditional on t is normally distributed with a mean of $\mu(t)$ with a standard deviation of σ .⁵ Given these assumptions, we can compute a likelihood function and hence, with the uniform prior, a posterior probability distribution.

4. CALIB and the “Scientists are not Bayesians” Argument. Having described CALIB, let us consider its bearing on the argument presented in section 2. Clearly, the above example and others like it show that summary generalizations to the effect that real scientists are not Bayesians are misleading at best, if not outright false. Nor is the present case an isolated occurrence, since the number of articles published in scientific journals that discuss applications of Bayesian methods has steadily increased within the last decade (Malakoff 1999, 1461). However, the argument can still proceed with the modified, and true, premise that Bayesian methods are a minority (i.e. (P1)). Let us grant this premise, but consider the relevance of CALIB to the inference from (P1) to the conclusion that scientists in general are committed to the principles of frequentist statistics (i.e., (C1)).

Most obviously, the use of CALIB calls into question whether scientists are committed to the frequentist view that the role of probabilities in statistics is not to

⁴ Interested readers can visit the site at <http://depts.washington.edu/qil/calib/>.

represent the degrees of confidence that one may reasonably have in hypotheses. CALIB produces a posterior probability distribution over a set of hypotheses concerning the calendar age of the item being dated. Suppose we generated an interval of .95 probability from such a posterior distribution. The interval and associated probability would assert that reasonable person could have a degree of confidence of .95 that the true date lies within the interval in question. The probabilities computed by CALIB in no way represent the relative frequency with which the program would, with high probability, produce estimates within a certain range of the correct result. In short, from the point of view of frequentist statistics, the probabilities in CALIB do not tell us what we want to know (cf. Mayo 1996, 414).

If scientists in general were committed to the principles of frequentist statistics, one would expect an uproar in the radiocarbon community once the workings of CALIB became known. However, no such backlash has ever occurred. Indeed, I found no mention of the difference between the use of probabilities in CALIB and the frequentist view of the role probabilities in statistical analysis. In defense of the “scientists are not Bayesians” argument, it could be observed the CALIB manual makes no mention of Bayesian statistics (Buck et al. 1996, 215; Stuiver and Reimer 1999). Nevertheless, it is not rare for the implicitly Bayesian character of CALIB, and similar programs, to be noted in the literature (cf. Bowman and Leese 1995, 99; Buck et al. 1991, 812).

Moreover, it is not difficult to explain why scientists might fail to notice that the probabilities in CALIB play a very different role than probabilities in classical statistics. That is, it is not rare for introductory statistics textbooks to misconstrue confidence

⁵ The standard deviation takes into account both the random error associated with the radiocarbon estimate and the uncertainty of the calibration curve (cf. Pearson 1987, 102; Stuiver and Reimer 1999, 1).

coefficients in frequentist methods as the degree of credence one may have in the hypothesis that the parameter lies within the interval. This mistake is most often found in introductory statistics textbooks written by scientists (who are not also statisticians) for students within their own field. For example, in *Statistical Methods for the Social Sciences*, we find the following definition, “The probability that the confidence interval contains the parameter is called the **confidence coefficient**” (Agresti and Finlay 1997, 125). In commenting on a subsequent example, the authors assert, “We can be 95% confident that this interval $[10.2 \pm .6]$ contains μ [the parameter being estimated]” (ibid. 127).⁶ Richard Drennan’s (1996), *Statistics for Archaeologists: A Commonsense Approach*, makes this interpretation of confidence coefficients its central theme. His “commonsense approach” amounts to dispensing with preset significance levels and interpreting confidence coefficients as degrees of credence that one may reasonably have in hypotheses (cf. ibid. vii; 125).

Given this common misinterpretation of confidence coefficients in introductory textbooks, it is hardly surprising that working scientists would fail to understand, much less be firmly committed to, the frequentist view of confidence intervals. Nor is it surprising that they would fail find anything untoward in a program like CALIB that computes a posterior probability distribution over a set of hypotheses. Therefore, one cannot justifiably infer a general commitment the frequentist view on the role of probabilities in statistical analysis from the widespread use of classical methods.

⁶ Cohen and Holiday (1996, 104-6) make precisely the same error. More advanced statistics texts seem less liable to make this mistake, however (cf. Lindgren 1993, 282-3).

5. More Thoroughly Bayesian Radiocarbon Calibration Programs. Although CALIB violates frequentist strictures through assigning probabilities to hypotheses, it is not clear that it violates the frequentist principle that estimators should be unbiased. This is because Bayes' theorem with a uniform prior probability distribution essentially behaves as a maximum likelihood estimator, and maximum likelihood estimators are generally unbiased. However, several researchers have advocated a more thorough going Bayesian approach to calibration in which archeological and historical information is allowed to influence the prior probability distribution (cf. Buck et al. 1991, Buck et al. 1996; Christen, Clymo, and Litton 1995; Goslar and Madry 1998).⁷ Moreover, some have created explicitly Bayesian calibration programs that can accommodate such information (cf. Ramsey 1995, 1998). Such approaches invariably use non-uniform prior probability distributions, making them biased by the standards of the classical theory of estimation.

The primary motivation for the development of such procedures has been the traditional Bayesian view that the prior probability distribution should accommodate prior information. As Buck et al. (1996, 3) put it: "Bayesian methods permit, indeed demand, that just such information (if it is available) be included." Christopher Ramsey argues that a uniform prior does not accurately represents the archeologist's beliefs even in cases in which little is known about the age of the item being dated, since "any traces of living matter are much more likely to be recent than they are to be extremely ancient, all other things being equal" (1998, 470-1).

Again, if scientists in general were firmly committed to the principles of frequentist statistics, one would expect that the radiocarbon community would condemn

⁷ Additional explicitly Bayesian statistical applications in radiocarbon calibration can be found by performing a key word search under "Bayesian AND Radiocarbon" in the Social Sciences Citation Index.

such Bayesian approaches to calibration. However, I found no mention in the literature of the fact that thoroughly Bayesian calibration methods would be biased estimators in the sense of the frequentist theory of estimation. On the contrary, the response to such Bayesian inroads into radiocarbon calibration and archeology in general so far seems mostly positive. In their review of developments in radiocarbon calibration, Bowman and Leese refer to the “emergent use of Bayesian statistics” (1995, 102). I located two reviews (Batt 1997, Scott 1997) of Buck et al.’s (1996), *Bayesian Approach to Interpreting Archaeological Data*, both of which were positive. One of these reviews states, “Bayesian analysis is becoming more and more popular and this book will encourage that interest” (Scott 1997, 219).

I found only one negative published reaction to the use of Bayesian statistics in radiocarbon calibration (Reece 1994), which was specifically directed at a particular essay (Buck et al. 1991). The Buck et al. article provides an example of how archeological information concerning phasing can be incorporated into the process of radiocarbon calibration via Bayes’ theorem. In his critique, Richard Reece argues against combining archeological information with radiocarbon dates in producing calendar date estimates, although his reasons for this view are not very clear. In any event, Reece’s objection is not that such a mixing of information will result in estimators that are biased in the frequentist sense, and the frequentist notion of an unbiased estimator is never mentioned. It is worth observing, furthermore, that some have proposed frequentist statistical methods for combining radiocarbon dates and archeological information in cases involving phasing (Orton 1983). There is no general prohibition in frequentist statistics on working information into the probability model of an estimation method, as

long as the resulting method is still unbiased, efficient, and so on. Reece's critique, therefore, seems to have little to do with the principles of frequentist statistics.

Despite the widespread use of CALIB and the increasing interest in more thoroughly Bayesian calibration programs, frequentist methods are commonly employed in radiocarbon research. To take just one case, it is not rare to see frequentist methods employed in the construction of the calibration curve. For example, Knox and McFadgen (1997) describe a procedure, based on the method of least squares, for smoothing out some of the curve's wiggles. However, when it comes to generating calendar dates from radiocarbon dates via a calibration curve assumed to be approximately accurate for the purposes of the analysis, Bayesian methods have a near monopoly. What is striking in all of this is the readiness of radiocarbon researchers to use frequentist methods in one context and Bayesian methods in another, with little or no regard for the conflicting principles underlying the two approaches. This pragmatically eclectic attitude towards statistical methods calls the "scientists are not Bayesians" argument into serious question. For it shows that the widespread use of a statistical method need not be accompanied by a steadfast commitment to the principles that underlie those methods.

6. Conclusion. The case described here shows that the charge that "scientists are not Bayesian" rests on an overly simplistic portrait of statistical practice in science. A better account of the matter is the following:

The statistician's toolbox today is more comprehensive than ever. Therein can be found Bayesian computational procedures alongside those based on other organizing principles such as maximum likelihood, maximum entropy, best

invariant and minimax estimation, and the most powerful specification of tests.

(Lad 1996, 4)

Bayesian methods are one of the statistical approaches available to the modern scientist, and the present case illustrates the willingness of scientists to use them when they perceive some advantage in doing so.

The present case also illustrates the difficulties involved inferring methodological principles accepted by scientists from observations of actual practice. It illustrates that the widespread use of a method M , for which P is the official justification, does not entail that scientists accept P and reject propositions inconsistent with it. Thus, the widespread use of frequentist methods does not provide a good basis for the conclusion that scientists in general reject the principles underlying Bayesian statistics.

REFERENCES

- Agresti, A. and B. Finlay (1997), *Statistical Methods for the Social Sciences*. 3rd edition. Upper Saddle River, NJ: Prentice Hall.
- Batt, Cathy (1997), "Book Review: Bayesian Approach to Interpreting Archaeological Data", *Antiquity* 71: 765-67.
- Bowman and Leese (1995), "Radiocarbon Calibration: Current Issues", *American Journal of Archaeology* 99: 102-5.
- Buck, C., Cavanagh B., Litton C. (1996), *Bayesian Approach to Interpreting Archaeological Data*. Chichester, England: Wiley & Sons.
- Buck, C., Kenworthy J., Litton C. (1991), "Combining archaeological and radiocarbon information: a Bayesian approach to calibration", *Antiquity* 65: 808-21.
- Campbell, B. and D. Loy (1996), *Humankind Emerging. Seventh edition*. New York, NY: Harper Collins.
- Christen, J. A., R. S. Clymo and C. D. Litton (1995), "A Bayesian Approach to the Use of ¹⁴C dates in the Estimation of the Age of Peat," *Radiocarbon* 37: 431-42.
- Cohen, L. and M. Holliday (1996), *Practical Statistics for Students*. London: Paul Chapman Publishing.
- Giere, R. (1988), *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.
- Drennan, R. (1996), *Statistics for Archaeologists: A Commonsense Approach*. New York: Plenum Press.
- Goslar, T. and W. Madry (1998), "Using the Bayesian Method to Study the Precision of Dating by Wiggle-Matching," *Radiocarbon* 40: 551-60.

- Howson, C. and P. Urbach (1993), *Scientific Reasoning: The Bayesian Approach*. La Salle, IL: Open Court.
- Knox, F. B. and B. G. McFadgen (1997), "Least-Squares Fitting a Smooth Curve to Radiocarbon Calibration Data," *Radiocarbon* 39: 193-204.
- Lad, Frank (1996), *Operational Subjective Statistical Methods*. New York: John Wiley and Sons.
- Lindgren, B. W. (1993), *Statistical Theory*. 4th edition. London: Chapman and Hall.
- Malakoff, David (1999), "Bayes Offers a 'New' Way to Make Sense of Numbers", *Science* 286: 1460-4.
- Mayo, Deborah (1996), *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.
- Naylor, J. and A. Smith (1988), "An Archaeological Inference Problem," *Journal of the American Statistical Association* 83: 588-95.
- Orton, C. (1983), "A statistical technique for integrating ¹⁴C dates with other forms of dating evidence," in J.G.B. Haigh (ed.), *Computer applications and quantitative methods in archaeology*. Bradford: University of Bradford Press, 115-24.
- Pazdur, M. and D. Michczynska (1989), "Improvement of the procedure for probabilistic calibration of radiocarbon dates," *Radiocarbon* 31: 824-32.
- Pearson, Gordon (1987), "How to Cope with Calibration," *Antiquity* 61: 98-103.
- Ramsey, Christopher B. (1995), "Radiocarbon Calibration and Analysis of Stratigraphy: The OxCal Program," *Radiocarbon* 37:425-30.
- _____ (1998), "Probability and Dating," *Radiocarbon* 40: 461-74.

- Reece, R. (1994), "Are Bayesian statistics useful to archeological reasoning?" *Antiquity* 68: 848-50.
- Scott, Marian (1997), "Book Review: Bayesian Approach to Interpreting Archaeological Data," *Radiocarbon* 39: 219.
- Stuiver, M. et al. (1998), "INTCAL98 Radiocarbon Age Calibration, 24,000-0 cal BP," *Radiocarbon* 40: 1041-84.
- Stuiver, M. and G. Pearson (1992), "Calibration of the Radiocarbon Time Scale, 2500-5000 BC," in R. E. Taylor, A. Long, and R. S. Kra (eds.), *Radiocarbon After Four Decades: An Interdisciplinary Perspective*. New York: Springer-Verlag, 19-33.
- _____ (1993), "High-precision bidecadal calibration of the radiocarbon time scale, AD 1950-500 BC and 2500-6000 BC," *Radiocarbon* 35: 1-23.
- Stuiver, M and H. Polach (1977), "Discussion: Reporting of ^{14}C data," *Radiocarbon* 19: 355-63.
- Stuiver, M. and P. Reimer (1986), "A computer program for radiocarbon age calibration," *Radiocarbon* 28: 1022-30.
- _____ (1999), *CALIB 4.1 Manual*. Available on line at:
<http://depts.washington.edu/qil/calib/>