Published in final edited form as:

Chemometr Intell Lab Syst. 2015 February 15; 141: 94–99. doi:10.1016/j.chemolab.2014.12.006.

A Bayesian approach to an interlaboratory comparison

George S. Jackson,

PRIME Laboratory, Purdue University, West Lafayette, IN 47907

Paul Muzikar, and

PRIME Laboratory, Purdue University, West Lafayette, IN 47907

Brent Goehring

Department of Earth and Environmental Sciences, Tulane University 70118

Abstract

Interlaboratory comparisons are an important check of the quality of a measurement technique. In this paper we examine the accelerator mass spectrometry (AMS) measurement of ^{41}Ca , an unstable isotope of calcium that has emerged as a valuable tracer for a variety of studies. We use a Bayesian framework to explore the quality and consistency of the AMS measurements made by Lawrence Livermore National Laboratory (LLNL) and the Purdue Rare Isotope Measurement Laboratory (PRIME Lab). This framework should be generalizable to other interlaboratory comparisons. The laboratories measured 47 samples, with each lab measuring an aliquot of each sample. The Bayesian approach allowed us to derive a probability distribution for four parameters reflecting the quality of the data, and to then address the following questions: (1) are the results from the two labs consistent? (2) are the uncertainties quoted by the two labs reasonable? We find that any consistent offset between the two labs is negligible, and that the uncertainties may be slightly underestimated.

Keywords

Bayesian; interlaboratory comparison; accelerator mass spectrometry

1 Introduction

Interlaboratory comparisons, in which aliquots of the same samples are measured at different laboratories and the results compared, serve many important purposes [1, 2, 3]. These purposes range from routine quality control to investigating problems arising in challenging new techniques. In a recent paper [4] we discussed such a comparison, in which PRIME Lab (the Purdue Rare Isotope Measurement Laboratory) and LLNL (Lawrence Livermore National Laboratory) performed ^{41}Ca accelerator mass spectrometry (AMS) [5]

phone: 765-494-3046; pmuzikar@purdue.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

measurements on a series of samples. These samples were not special, standard material; rather they were unknowns that had already been chemically prepared and measured at PRIME Lab for other studies [6, 7]. Thus, the results are representative of the actual quality of the measurements being done at both labs. In our paper [4], the results from the two labs were compared in a variety of ways; one particular method was grounded in a Bayesian approach. In this new paper we focus on the Bayesian approach [8, 9, 10], expanding on our original discussion. We analyze more aspects of the data, and explore more possibilities that are inherent in the Bayesian methodology. We hope in this paper to both provide a more in-depth look at our particular data set, and more generally, to provide ideas useful in a wider context.

^{41}Ca is an unstable isotope of calcium, with a half-life of 1.03×10^5 years. This half-life is short enough to make ^{41}Ca rare in the environment, but long enough to make decay counting impractical. It is an ideal candidate for AMS studies, since AMS has an extremely low detection limit [5]. Its AMS measurement is reported as the following ratio: $R = ^{41}\text{Ca}/\text{Ca}$, where the numerator is the number of ^{41}Ca atoms in the sample, and the denominator is the total number of the stable isotopes of calcium (mainly ^{40}Ca) in the sample.

^{41}Ca has emerged as a valuable tool for a variety of studies. For example, ^{41}Ca has been used to glean information about pre-atmospheric meteorite dimensions [11, 12]. For biomedical studies, the low detection limit and the long half-life mean that a single, safe dose can be given to an animal or a human subject, and then research can be conducted for the biological lifetime of the subject without another dose ever being administered [13]. In the 1990's it was suggested that ^{41}Ca could be used as a tracer to elucidate calcium metabolism in humans [14, 15], and this has now emerged as the most important application of this nuclide [16, 17, 18, 19]. For example, clinical trials have studied the efficacy of both commercial anti-osteoporosis drugs and of botanical treatments [6, 7]. The increasing use of ^{41}Ca in these biomedical studies adds importance to our interlaboratory comparison.

The technical challenges of AMS ^{41}Ca measurements mostly stem from the low ion source currents that are produced when CaF_2 is used as the target material, and from the large number of interfering ions that are present in the detector when measuring ^{41}Ca . The easily prepared compound CaF_2 is the preferred target material for studies that require high sample throughput.

The data set we will consider is a set of measurements on 47 samples, in which R ranged from about 10^{-10} to about 36×10^{-10} . (The data are presented in [4].) PRIME Lab and LLNL separately measured an aliquot of each sample, and reported an R value and an uncertainty for each sample. As already mentioned, these samples were not standards, so the true values are not known. We will use a Bayesian approach to address the following questions: (1) are the results from the two laboratories consistent? (2) are the uncertainties quoted by the two labs reasonable? Since the samples were all unknowns, the answer to a third question, ((3) how close are the results to the true values?), involves quantitative input that goes beyond the data; this input is incorporated in the prior distribution, as we discuss.

2 Theoretical development

2.1 Bayesian approach

In this section, we develop the formalism we will use. We enlarge on the Bayesian discussion in our previous paper[4], again following the approach presented by Dose [10] (see also [8, 9]). A closely related problem arises when a group of independent laboratories measures the same quantity, and we want to combine the results to determine the best estimate. For approaches related to ours, see [20, 21, 22].

Let $[x_i, \sigma_{Li}]$ be the LLNL data points and their associated uncertainties for sample i , and $[y_i, \sigma_{Pi}]$ be the PRIME Lab results for sample i . So, for example, LLNL asserts that for sample i the probability density for the true value has an average equal to x_i and a standard deviation of σ_{Li} . We will define four parameters, and use a Bayesian approach to formulate a probability distribution for them. We will use, throughout this paper, the notation $p(A|B)$, which means the probability of A , given B .

First, we allow for a consistent offset of $-a$ of the LLNL samples from the true value, such that $x_i + a$ is the true value. We define a similar consistent offset for the PRIME Lab samples, such that $y_i + b$ is the true value. We will see what the data allow us to say about these offsets. We note that other choices are possible; for example, we could assume that the LLNL offset is proportional to x_i , so that the true value is $x_i + \gamma x_i$, where γ is the same for all of the LLNL samples. In this paper we will focus our study on the offsets a and b .

Next, we also allow for the fact that the labs may have misestimated their uncertainties. A simple model for this is that $v\sigma_{Li}$ and $w\sigma_{Pi}$ are the true uncertainties of LLNL and PRIME Lab. This defines the two non-negative parameters v and w . For example, if $v < 1$, this means that LLNL is overestimating the uncertainty it assigns to its results, while if $v > 1$ the uncertainties are being underestimated.

So, we have defined four parameters which characterize the quality of the data; a and b tell us how accurate the measurements are, while v and w tell us if the quoted uncertainties are correct. We should emphasize that our approach is not based on a more fundamental, detailed model of the experimental situation. Rather, these four parameters are a simple way to allow for a range of errors in the reported results. As mentioned above, other choices are possible; we hope this paper makes clear how alternative models could be handled in a Bayesian spirit.

In the Bayesian approach, we do not simply use the data to estimate the values of these parameters. Instead, the data are used to generate a probability distribution (since these are continuous parameters, it is in fact a probability density function, or pdf) for these parameters. This is a more informative result, since from the complete pdf we can compute all averages (such as $\langle a \rangle$, $\langle a^2 \rangle$, ...), generate contour plots, and calculate probability distributions for any subset of our parameters.

To derive our basic equation, we start with Bayes theorem:

$$p(a, b, v, w|data)p(data) = p(data|a, b, v, w)p(a, b, v, w) \quad (1)$$

We first discuss the meaning of each of the four probability densities in the previous equation. $p(a, b, v, w|data)$ is the joint probability density for our four unknowns (a, b, v, w) given the data. This function is the result we want. $p(data)$ is the probability of acquiring the data; it does not depend on the four unknowns, and can be dropped if we are willing to normalize our result. Thus, we write

$$p(a, b, v, w|data) = c p(data|a, b, v, w)p(a, b, v, w) \quad (2)$$

The value of c will be determined by requiring that

$$\int da \int db \int dv \int dw p(a, b, v, w|data) = 1 \quad (3)$$

Next we discuss $p(data|a, b, v, w)$; this is the probability of obtaining our data, given the values of a, b, v, w , and is also known as the likelihood. We must use our knowledge of the measurement process to construct an explicit expression for this likelihood.

Finally, we come to the pdf $p(a, b, v, w)$. This is called the prior, and embodies our information about the values of these parameters before we have made our measurements. Thus, a nice way to view equation (2) is as follows. The prior, $p(a, b, v, w)$ represents our knowledge of the four parameters before we collected the AMS data. This knowledge is then updated through the likelihood, in order to produce $p(a, b, v, w|data)$, which represents our improved state of knowledge.

2.2 Derivation of the likelihood function

To derive the likelihood function, we start with

$$p(data|a, b, v, w) = \prod_i p(x_i, y_i|a, b, v, w) \quad (4)$$

where $\{x_i\}$ are the R values measured at LLNL, and $\{y_i\}$ are the R values measured at PRIME Lab. Then, for each sample we may write

$$p(x_i, y_i|a, b, v, w) = \int dz_i p(x_i, y_i|z_i, a, b, v, w)p(z_i|a, b, v, w) \quad (5)$$

where z_i is the (unknown) true R value for sample i .

We must now make several more definite assumptions. We take the measurements to have a Gaussian distribution about the true value, with the shifts a and b included:

$$p(x_i, y_i | z_i, a, b, v, w) \propto \left(\frac{\exp\left(-\frac{(x_i + a - z_i)^2}{2v^2\sigma_{Li}^2}\right)}{v\sigma_{Li}} \right) \left(\frac{\exp\left(-\frac{(y_i + b - z_i)^2}{2w^2\sigma_{Pi}^2}\right)}{w\sigma_{Pi}} \right) \quad (6)$$

Note that the parameters v and w each appear in two places, since the overall normalization of each Gaussian factor involves its uncertainty. Numerical factors that do not involve the parameters a , b , v and w may be omitted, since we will normalize at the end. The Gaussians are a reasonable choice, and are convenient for evaluating the integral. We take the conditional probability $p(z_i | a, b, v, w)$ to be a constant, since the true value z_i should not depend on these parameters, and will have a very broad range of possible values.

We may then evaluate the integral in equation (5) to obtain:

$$p(x_i, y_i | a, b, v, w) \propto (v^2\sigma_{Li}^2 + w^2\sigma_{Pi}^2)^{-1/2} \exp\left(-\frac{(x_i - y_i + a - b)^2}{2(v^2\sigma_{Li}^2 + w^2\sigma_{Pi}^2)}\right) \quad (7)$$

We note that the likelihood involves a and b only in the combination $a - b$. Thus, instead of using a and b , it is instructive to make a change of variables:

$$r = a - b, Q = (a + b)/2 \quad (8)$$

We then write our final form for the likelihood as

$$p(\text{data} | r, Q, v, w) = \prod_i (v^2\sigma_{Li}^2 + w^2\sigma_{Pi}^2)^{-1/2} \exp\left(-\frac{(x_i - y_i + r)^2}{2(v^2\sigma_{Li}^2 + w^2\sigma_{Pi}^2)}\right) \quad (9)$$

2.3 Discussion of the prior

To formulate the prior can be a somewhat tricky undertaking, since we are forced to turn our rather vague prior knowledge into an explicit mathematical form. For some applications, the goal is to make the prior as noncommittal as possible, while in other cases some definite information can be included. For our case, we reason as follows. We give a and b the same distributions, since we make no assumptions about the relative accuracies of LLNL and PRIME Lab. For the same reason we give v the same prior distribution as w .

We assume that positive and negative values of a and b are equally likely; a simple model is then a Gaussian distribution for a and for b , each centered on zero. Thus we write

$$p(a, b, v, w) \propto \left(\exp\left(-\frac{a^2}{2m^2}\right) \right) \left(\exp\left(-\frac{b^2}{2m^2}\right) \right) f(v, w) \quad (10)$$

The value of m reflects how much we may reasonably expect the measured AMS values for R to systematically deviate from the true values.

For the parameters v and w , we expect that neither lab is seriously overestimating its uncertainties. Thus, a reasonable model for the function $f(v, w)$ is

$$f(v, w) \propto H(v - u_0)H(w - u_0) \quad (11)$$

where $H(x)$ is the Heaviside function, which is zero for $x < 0$ and one for $x > 0$. We may chose $u_0 = 1/2$, which means we think it is very improbable that the true uncertainty is less than half the quoted value for either lab. Our final form for the prior is then given by

$$p(r, Q, v, w) \propto \left(\exp\left(-\frac{r^2}{4m^2}\right) \right) \left(\exp\left(-\frac{Q^2}{m^2}\right) \right) H(v - u_0)H(w - u_0) \quad (12)$$

where we have shifted to the variables r and Q . An important feature of the Bayesian approach is that we can check to see if our results are sensitive to changes in the prior. If two different, equally plausible priors produce significantly different results, then we can conclude that our data are not definitive enough to answer the question we have posed. For example, in our case we may calculate our result, $p(a, b, v, w|data)$, for different reasonable values of m and u_0 , and examine the changes in this pdf.

2.4 Combined result

We now combine the results from the last two subsections. This gives

$$p(r, Q, v, w|data) = c p(data|r, Q, v, w) p(r, Q, v, w) \quad (13)$$

where we use equations (9) and (12) on the right hand side of (13). The coefficient c is selected to normalize the final result:

$$\int dr \int dQ \int dv \int dw p(r, Q, v, w|data) = 1 \quad (14)$$

To understand the implications of equation (13), we note that the likelihood can directly reflect the value of r , since if LLNL and PRIME Lab are systematically off by different amounts, this will show up in the data; conversely, if both labs are off by the same amount, this will not show up. In addition, the product of the Gaussian priors for a and b has the special property that, when we transform to r and Q , we obtain a function of r times a

function of Q . So the upshot is that the data have no effect on our knowledge of Q : the information about Q contained in the prior is not updated by the data.

Note that if the prior contained correlations between r and Q , the situation would be different. Even though the likelihood does not contain Q , an updated knowledge for r would lead to an updated knowledge for Q through the correlations in the prior.

3 Results

To use equation (13), we plug in the $[x_i, \sigma_{Li}], [y_i, \sigma_{Pi}]$ from the 47 samples [4]. These numbers are all dimensionless, and we will use units of 10^{-10} for all of them; since the $^{41}\text{Ca}/\text{Ca}$ ratios are so small, it is handy to omit the 10^{-10} factor. Thus, x_i and y_i span about 1 – 36, and the uncertainties span about 0.02 – 2.0. So, a , b , r , m and Q are reported in units of 10^{-10} . On the other hand, the multiplicative parameters v and w are expected to be of order one, and so no rescaling is needed.

In these units, we chose $m = 1$. This choice for m is somewhat pessimistic, since the accuracy of the AMS measurements is almost certainly better than this; we will later examine the consequences of a smaller value for m . As discussed in a previous section, we choose $u_0 = 1/2$. This choice ($m = 1$, $u_0 = 1/2$) will be our standard, default choice for the prior.

3.1 Results for r, a and b

To generate a probability distribution for r , we integrate out the other parameters:

$$P(r|data) = \int dQ \int dv \int dw p(r, Q, v, w|data) \quad (15)$$

In Figure 1 we show the resulting distribution for r . We can also compute the average and standard deviation for r ; we obtain the following (recall these numbers are in units of 10^{-10}):

$$\langle r \rangle = 0.014\sigma_r = 0.017 \quad (16)$$

Thus, for the range of values measured (between about 1 and 36) the LLNL and PRIME Lab results agree quite well. Any consistent offset between the two labs is most likely in the range $-0.03 - +0.06$. In addition, the average value of r is within one sigma of zero. As already mentioned, the pdf for Q is unaffected by the data, so we have $\langle Q \rangle = 0$ and $\sigma_Q^2 = m^2/2 = 1/2$.

The results for r and Q can be converted into results for a and b . In particular, we have

$$\begin{aligned} \langle a \rangle &= \frac{1}{2} \langle r \rangle + \langle Q \rangle = \frac{1}{2} \langle r \rangle \quad \langle b \rangle = -\frac{1}{2} \langle r \rangle + \langle Q \rangle = -\frac{1}{2} \langle r \rangle \\ (17) \end{aligned}$$

$$\sigma_a^2 = \sigma_b^2 = \sigma_r^2/4 + \sigma_Q^2 = \sigma_r^2/4 + m^2/2 \quad (18)$$

With our standard choice ($m = 1.0$ and $u_0 = 1/2$) we find

$$\langle a \rangle = -\langle b \rangle = 0.007\sigma_a = \sigma_b = 0.71 \quad (19)$$

To see how the data have improved our knowledge of a and b , we note that if we simply use the prior (12) to do the calculation, we obtain:

$$\langle a \rangle = \langle b \rangle = 0\sigma_a = \sigma_b = 1.0 \quad (20)$$

Thus, we see that the data have reduced the uncertainties on both parameters from 1.0 to 0.71. If we take into account the very small values given to $\langle a \rangle$ and $\langle b \rangle$ by the data, we see that overall, the data let us infer smaller possible values for the magnitudes of these offsets.

This point deserves further discussion. By looking at equation (18), we can see that the uncertainty in a (or in b) has two contributions. The contribution from σ_r depends on the data, and can be made small if the two labs are doing a good job. Conversely, the contribution from σ_Q is not affected by the data; for a given choice of m , this provides a lower limit on σ_a . We can understand this by recalling that $-a$ is the systematic deviation of the LLNL values from the true values; since we do not know the true values, the interlab comparison can only go so far in tightening our knowledge of a .

3.2 Results for v and w

In this section we present our results for the parameters v and w . To derive the probability distribution for them, we integrate out r and Q :

$$p(v, w | data) = \int dr \int dQ p(r, Q, v, w | data) \quad (21)$$

In Figure 2 we show contours of equal probability density in the $v - w$ plane. For this plot, we have omitted the Heaviside functions from the prior (equation (12)); thus, all positive values of v and w were allowed, and we can see that there is nontrivial statistical weight in

the region $v < 1/2$. In Figure 3 we show plots of the pdf's for v and for w ; in these plots the Heaviside functions are used.

If we compute the average and standard deviation for these parameters we obtain the following

$$\langle v \rangle = 1.38\sigma_v = 0.62 \quad (22)$$

$$\langle w \rangle = 1.34\sigma_w = 0.21 \quad (23)$$

The fact that $\langle v \rangle$ and $\langle w \rangle$ are higher than one is evidence that the true uncertainties are slightly larger than the quoted values; however, the large spread in the likely values means that $\langle v \rangle$ is within one σ of one, and that $\langle w \rangle$ is within two σ of one.

3.3 Sensitivity of the the results to the prior

We have presented results for our parameters, using the values $m = 1$ and $u_0 = 1/2$ in the prior. We now investigate how much these results change if we use different, but still reasonable values for m and for u_0 .

First, we note that the probability distribution for r has significant statistical weight only in the range $-0.03 - +0.06$. This range covers values much smaller than $m = 1$. So it is clear that choosing values of m larger than 1 will not change our results for r , v and w . A calculation using $m = 0.1$ produces results for r , v and w that do not differ significantly from the $m = 1.0$ results. If we do a calculation with $m = 0.01$, we find that v and w are not significantly altered, but that $\langle r \rangle$ is a little smaller ($\langle r \rangle = 0.006$, $\sigma_r = 0.01$).

Next, we consider changing the value of u_0 ; recall that in setting $u_0 = 1/2$ we rule out the possibility that the true uncertainties are less than $1/2$ of the quoted values. If we look at figure 3, we see that v , the LLNL multiplier, would have had substantial statistical weight for values below $v = 0.5$, if the prior with $u_0 = 1/2$ had not left out this weight. So, we try a prior with $u_0 = 0.1$, which allows the true uncertainty for each lab to be as small as $1/10$ of the claimed value. This change in the prior does lead to somewhat different results. Using $m = 1.0$ and $u_0 = 0.1$, we obtain

$$\langle r \rangle = 0.014\sigma_r = 0.016 \quad (24)$$

$$\langle v \rangle = 1.18\sigma_v = 0.70 \quad (25)$$

$$\langle w \rangle = 1.37\sigma_w = 0.21 \quad (26)$$

We see that allowing both v and w to be as small as 0.1 has a much bigger effect on the LLNL multiplier v than on the PRIME Lab multiplier w . The pdf of equation (13) will assign significant statistical weight to small values of v , if the prior does not forbid this.

3.4 Non-Bayesian approaches

In this section we provide a brief comparison of our results to those obtained with more traditional methods. First we consider the parameter r , the offset between the two labs. Equation (16), along with Figure 1, provides our basic result: $\langle r \rangle = 0.014 \pm 0.017$. In our previous paper, we applied the Wilcoxon rank test, which saw no significant difference between the two labs. Thus, this test is consistent with our Bayesian conclusion. For another check on r , we reason as follows. In section 2.1 we defined, for sample i , $x_i = z_i - a$ and $y_i = z_i - b$, where z_i is the true value, x_i is the LLNL value, and y_i is the PRIME Lab value. Eliminating z_i gives $y_i = x_i + a - b = x_i + r$. Thus, we performed a least-squares fit of a line (with slope of one) to our data points; the y-intercept should then give an estimate for the value of r . We obtained $r = 0.014 \pm 0.015$, which agrees very well with the Bayesian answer.

Next, we consider a simple check on whether the labs have assigned their uncertainties accurately. To do this, we assume that there is no consistent offset between the two labs. We then define

$$\chi^2 = \sum_i \frac{(x_i - y_i)^2}{\sigma_{Li}^2 + \sigma_{Pi}^2} \quad (27)$$

If we evaluate this sum for all 47 samples, we obtain $\chi^2 = 79.7$. With 47 degrees of freedom, the probability of obtaining a value this large is 0.002. Thus, we have reason to believe that the uncertainties are underestimated.

4 Discussion

In this paper we have shown how a Bayesian approach can be applied to an interlaboratory comparison. The basic strategy is to define a set of parameters which characterize the quality of the data (in our case a , b , v and w) and then use Bayes theorem (equation (1)) to derive a probability distribution for these parameters. This final result, which we wrote as $p(a, b, v, w | data)$, or as $p(r, Q, v, w | data)$, depends on the data collected for the interlab comparison, and on the prior. The prior is a pdf which reflects our knowledge of the parameters before we use the interlab comparison data.

In our case, we concluded that the magnitude of any consistent offset between the LLNL and the PRIME Lab R values was no bigger than about 0.05, or about 5×10^{-12} in the original, unscaled units. In addition, even though the true R values of the samples are not known, we were able to narrow down the uncertainties on a and b , where $-a$ is the systematic offset of

the LLNL values from the true value, and $-b$ is the analogous offset for PRIME Lab. It is important to note, however, that the uncertainties for a and b also involve a contribution which depends on our judgement, and is not affected by the data we used in this study.

We also assessed whether the two labs were giving accurate estimates of their uncertainties. Our results indicate that the PRIME Lab uncertainties should be rescaled by a factor somewhere between about 0.9 and 1.7, while the LLNL uncertainties should be rescaled by a factor ranging between 0.2 and 2.6. While these ranges are too broad to be very informative, they do serve to indicate that neither lab is off by an order of magnitude in estimating its uncertainty, and that a value close to one for the scaling factor is not ruled out for either lab.

Acknowledgements

We would like to thank Darren Hillegonds for arranging the measurements at LLNL. This work was supported by National Science Foundation grant EAR 0112480 and NIH grants P50-ATT00477 and R21-CA127671.

References

1. Vander Heyden Y, Smeyers-Verbeke J. Set-up and evaluation of interlaboratory studies. *J. Chromatogr. A*. 2007; 1158:158–167. [PubMed: 17339041]
2. Bryant C, Carmi I, Cook G, Gulliksen S, Harkness D, Heinemeier J, McGee E, Naysmith P, Possnert G, Scott M, van der Plicht H, van Strydonck M. Sample requirements and design of an inter-laboratory trial for radiocarbon laboratories. *Nucl. Instrum. Meth. B*. 2000; 172:355–358.
3. Martin GJ, Koziat J, Rossmann A, Dennis J. Site-specific natural isotope fractionation in fruit juices determined by deuterium NMR: an European inter-laboratory comparison study. *Anal. Chim. Acta*. 1996; 321:137–146.
4. Jackson GS, Hillegonds DJ, Muzikar P, Goehring B. Ultra-trace analysis of ^{41}Ca in urine by accelerator mass spectrometry: an inter-laboratory comparison. *Nucl. Instrum. Meth. B*. 2013; 313:14–20.
5. Tuniz, C., Bird, JR., Fink, D., Herzog, GF. Accelerator Mass Spectrometry: Ultrasensitive Analysis for Global Science. CRC Press; 1998.
6. Cheong JMK, Martin BR, Jackson GS, Elmore D, McCabe GP, Nolan JR, Barnes S, Peacock M, Weaver CM. Soy isoflavones do not affect bone resorption in postmenopausal women: A dose-response study using a novel approach with Ca-41. *J. Clin. Endocr. Metab*. 2007; 92:577–582. [PubMed: 17148567]
7. Weaver CM, Martin BR, Jackson GS, McCabe GP, Nolan JR, McCabe LD, Barnes S, Reinwald S, Boris ME, Peacock M. Antiresorptive Effects of Phytoestrogen Supplements Compared with Estradiol or Risedronate in Postmenopausal Women Using Ca-41 Methodology. *J. Clin. Endocr. Metab*. 2009; 94:3798–3805. [PubMed: 19584189]
8. D'Agostini G. Bayesian inference in processing experimental data: principles and basic applications. *Rep. Prog. Phys*. 2003; 66:1383–1419.
9. D'Agostini G. Bayesian reasoning in data analysis. World Scientific. 2003
10. Dose V. Bayesian inference in physics: case studies. *Rep. Prog. Phys*. 2003; 66:1421–1461.
11. Welten KC, Caffee MW, Hillegonds DJ, McCoy TJ, Masarik J, Nishiizumi K. Cosmogenic radionuclides in L5 and LL5 chondrites from Queen Alexandra Range, Antarctica: Identification of a large L/LL5 chondrite shower with a preatmospheric mass of approximately 50,000 kg. *Meteorit. Planet. Sci*. 2011; 46
12. Herzog GF, Albrecht A, Ma PX, Fink D, Klein J, Middleton R, Bogard DD, Nyquist LE, Shih CY, Garrison DH, Reese Y, Masarik J, Reedy RC, Rugel G, Faestermann T, Korschinek G. Cosmic-ray exposure history of the Norton County enstatite achondrite. *Meteorit. Planet. Sci*. 2011; 46:284–310.

13. Jackson GS, Weaver C, Elmore D. Use of accelerator mass spectrometry for studies in nutrition. *Nutr. Res. Rev.* 2001; 14:317–334. [PubMed: 19087429]
14. Elmore D, Bhattacharyya MH, Sacco-Gibson N, Peterson DP. Calcium-41 as a long-term biological tracer for bone resorption. *Nucl. Instrum. Meth. B.* 1990; 52:531–535.
15. Johnson RR, Berkovits D, Boaretto E, Gelbart Z, Ghelberg S, Meirav O, Paul M, Prior J, Sossi V, Venczel E. Calcium resorption from bone in a human studied by ^{41}Ca tracing. *Nucl. Instrum. Meth. B.* 1994; 92:483–488.
16. Lee WH, Wastney ME, Jackson GS, Martin BR, Weaver CM. Interpretation of Ca-41 data using compartmental modeling in post-menopausal women. *Anal. Bioanal. Chem.* 2011; 399:1613–1622. [PubMed: 21152905]
17. Freeman SPHT, King JC, Vieira NE, Woodhouse LR, Yergey AL. Human calcium metabolism including bone resorption measured with ^{41}Ca tracer. *Nucl. Instrum. Meth. B.* 1997; 123:266–270.
18. Denk E, Hillegonds D, Vogel J, Synal A, Geppert C, Wendt K, Fattinger K, Hennessy C, Berglund M, Hurrell RF, Walczyk T. Labeling the human skeleton with Ca-41 to assess changes in bone calcium metabolism. *Anal. Bioanal. Chem.* 2006; 386:1587–1602. [PubMed: 17033771]
19. Sharma M, Bajzer Z, Hui SK. Development of Ca-41-Based Pharmacokinetic Model for the Study of Bone Remodelling in Humans. *Clin. Pharmacokinet.* 2011; 50:191–199. [PubMed: 21294596]
20. Toman B. Bayesian approaches to calculating a reference value in key comparison experiments. *Technometrics.* 2007; 49:81–87.
21. Shirono K, Tanaka H, Ehara K. Bayesian statistics for determination of the reference value and degree of equivalence of inconsistent comparison data. *Metrologia.* 2010; 47:444–452.
22. Burr T, Williams B, Croft S, White M, Hanson K. Meta-analysis options for inconsistent nuclear measurements. *Nucl. Sci. Eng.* 2013; 173:15–27.

Highlights

- Both PRIME Lab and LLNL performed ^{41}Ca AMS measurements on a set of 47 samples.
- We used a Bayesian framework to compare the results of the two labs.
- We found that any offset between the two labs is small.
- We also were able to examine whether the labs estimated their uncertainties accurately.
- We analyzed how sensitive our results were to the prior.

Figure 1.

This shows the normalized probability density for $r = a - b$, computed using equation (15). Note that units of 10^{-10} are used for r .

Figure 2.

This shows contour lines of equal probability density in the $v - w$ plane. As mentioned in the text, the Heaviside functions in the prior were omitted in generating this plot, so that all positive values of v and w were allowed. The (unnormalized) values for the probability density on the four contour lines shown are 0.21, 0.021, 0.0021, and 0.00021.

Figure 3.

(a) This shows the normalized pdf for v , the factor which rescales the LLNL uncertainties. We can see that the prior, which ruled out values of v less than 0.5, has removed significant statistical weight at low values of v . (b) This shows the normalized pdf for w , the factor which rescales the PRIME Lab uncertainties. In this case, there is very little statistical weight at values less than 0.5 for the prior to remove.