Published in final edited form as:

J Anal At Spectrom. 2012 ; 27(4): 595–603. doi:10.1039/C2JA10369A.

An interlaboratory comparison of bone lead measurements via K-shell X-ray fluorescence spectrometry: validation against inductively coupled plasma mass spectrometry

David J. Bellis^a, Andrew C. Todd^b, and Patrick J. Parsons^{a,c}

^aLaboratory of Inorganic and Nuclear Chemistry, Wadsworth Center, New York State Department of Health, Albany, NY, 12201, USA

^bDepartment of Preventive Medicine, Mount Sinai School of Medicine, New York, NY, 10029, USA

^cDepartment of Environmental Health Sciences, School of Public Health, The University at Albany, State University of New York, Albany, NY, 12201, USA

Abstract

¹⁰⁹Cd-based K-shell X-ray fluorescence spectrometry (hereafter, for brevity, XRF) is used, often in epidemiological studies, to perform non-invasive, *in vivo* measurements of lead in bone. We conducted the first interlaboratory study of XRF *via* the circulation of nine goat tibiae in which the mean lead value ranged from 4.0 $\mu\text{g g}^{-1}$ to 55.3 $\mu\text{g g}^{-1}$ bone mineral. The test tibiae were subsequently analyzed *via* nitric acid digestion followed by lead determination by inductively coupled plasma mass spectrometry (ICP-MS) – along with certified reference materials for bone lead – thus providing measurement traceability to SI units. Analysis of dried bone for lead *via* nitric acid digestion and ICP-MS yields mass fraction data in units of $\mu\text{g g}^{-1}$ dry weight. The mean bone lead value based on ICP-MS analysis ranged from 1.8 $\mu\text{g g}^{-1}$ to 35.8 $\mu\text{g g}^{-1}$ dry weight. For comparison purposes, XRF-measured Pb values ($\mu\text{g g}^{-1}$ bone mineral) were converted into the ICP-MS-measured units ($\mu\text{g g}^{-1}$ dry weight bone) by multiplying the former by the average ash fraction from the nine tibiae. Eight of the XRF systems did not yield a significant bias for any of the nine tibiae; one system was biased for one of the tibiae; two systems were biased for two tibiae; one system was biased for four tibiae; two systems (813-1 and 804-2) were biased for five tibiae and one system (801-1) was biased for six of the nine tibiae. Average bias for the systems (under those particular operating conditions) that were biased for the majority of samples ranged from $-2.6 \mu\text{g g}^{-1}$ (-15.7%) to $5.1 \mu\text{g g}^{-1}$ (30.7%) dry weight bone. All participants now have the ICP-MS data, allowing any corrective actions deemed necessary to be implemented. The ICP-MS data, however, indicated that the lead mass fraction varied considerably with the sampling location within the tibiae, to the extent of exceeding XRF variability for the higher lead values. Material heterogeneity is an unavoidable reality of measuring lead in bone.

1. Introduction

¹⁰⁹Cd-based, energy dispersive, K-shell X-ray fluorescence measurements of lead in bone (hereafter, for brevity, XRF) have been widely employed, *in vivo*, in epidemiological studies to provide a non-invasive assessment of long-term exposure to lead.^{1–4} Recently, we conducted the first interlaboratory study of XRF methods by circulating a set of nine goat

tibiae to eleven laboratories, five of which measured the tibiae with each of two measurement systems.⁵ The tibiae were collected from goats that had been dosed with lead over an extended period of time in order to produce blood pools physiologically enriched in lead, at concentrations of relevance for both environmental and occupational human exposure levels, for use in the New York State Department of Health's Blood Lead Proficiency Testing Program. The bones from these animals were consequently enriched in lead at levels, *i.e.*, mass fractions, similar to those that have been observed for humans ($2 \mu\text{g g}^{-1}$ to $60 \mu\text{g g}^{-1}$ dry weight, or $3 \mu\text{g g}^{-1}$ to $94 \mu\text{g g}^{-1}$ bone mineral), the latter using the conversion factor of 1.5681 inferred from Woodard⁶ and treating, as have others,^{7,8} 'ashed bone' and 'bone mineral' as synonymous). The ultimate goal of the XRF interlaboratory/system agreement study was the standardization of XRF bone lead measurements. The achievement of this goal requires the comparison between the XRF data and the SI-traceable ICP-MS-measured values of the same tibiae reported herein.

Current XRF systems use the 88.034 keV γ -rays from ^{109}Cd to fluoresce lead K-shell x rays that are then detected *via* a germanium detector and its associated spectroscopy electronics. Although individual protocols vary, calibration is generally performed *via* measurement of plaster-of-Paris ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) calibration standards ('phantom' bones) doped with lead. Lead X-ray intensities are normalized to the intensity of the XRF spectrum peak arising from coherent scatter (originating from the 88.035 keV ^{109}Cd γ -rays that are used to fluoresce the K-shell electrons of lead), because the X-ray-to-coherent peak ratio is relatively unaffected by practical considerations such as overlying tissue thickness, source-to-sample distance and bone size and shape.^{8,9} A matrix conversion factor (which is also a function of photon scattering angle) accounts for the different amounts of coherent scatter yielded by the matrices of the calibration standard and the unknown sample (assumed to be calcium hydroxyapatite, which has the general chemical formula $\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$, for the *in vivo* case).

Some matrix conversion factors have been reported in the literature. Somervaille *et al.* (1985) reported a plaster-to-ashed-bone conversion factor of 1.4565 for a mean scattering angle of 153.25° (they received "some dried and some ashed samples of a tibia section" but do not explicitly state a factor for converting from plaster to dry weight bone).⁸ Other researchers¹⁰ have used a plaster-to-bone-mineral conversion factor of 1.462 for a mean scattering angle of 160° (a value recalculated to be 1.4602 by Todd⁹), and 1.4615 for plaster-to-bone-mineral at a mean scattering angle 167.7° ,¹¹ (and later reproduced by Todd as 1.46168).⁹

Four previous reports in the literature have compared measurements of lead in human bone obtained *via* both ^{109}Cd -based K-shell XRF and independent analytical methods such as atomic absorption spectrometry (AAS) and ICP-MS.^{12–15} In one study, 30 tibia and metatarsal samples from 30 subjects yielded no evidence of a statistically significant difference between XRF and AAS measurements.¹⁵ Other researchers measured lead in three legs at eight locations *via* both XRF and AAS; the agreement was reported as a correlation coefficient of 0.98.¹⁴ In a more recent report from the same laboratory, ICP-MS was used to measure lead concentrations in eight cadaver legs, and "good" agreement was reported in terms of correlation coefficients.¹³ Previously, we compared XRF-measured lead (in human cadaver bones), in units of $\mu\text{g g}^{-1}$ plaster, to AAS-measured lead following wet ashing in concentrated HNO_3 , in units of $\mu\text{g g}^{-1}$ dry weight.¹² For the interlaboratory study described here, we requested that participants report bone lead data to us in units of $\mu\text{g g}^{-1}$ bone mineral.

The principal aim of the present work was to assess the accuracy of the XRF-measured bone lead data obtained as part of the interlaboratory study, and to make an initial assessment of

performance. Herein, we report the ICP-MS-measured lead content of the intact bare tibiae used for the interlaboratory study; assess the accuracy of the XRF data from each measurement system; and begin to address bias correction. We also include data from one laboratory that was not included in the initial report of the interlaboratory study.⁵

2. Materials and methods

2.1. Source of the goat tibiae

A full description of the nine goat tibiae has been given elsewhere.⁵ Briefly, the tibiae were collected *post mortem* from goats (*Capra hircus*) that are maintained by the New York State Department of Health (NYSDOH) for periodic dosing with lead acetate, in order to produce blood endogenously enriched in lead for use in the NYS DOH Blood Lead Proficiency Testing (PT) program.¹⁶ Periodically, goats are euthanized on the recommendation of the facility veterinarian (no animals were euthanized for the purposes of this study); others die of natural causes. Over time, the bones from many such goats have been harvested and archived at -80°C for research purposes. The nine goats used for the study were selected from a pool of more than 50 to provide a substantial range in cumulative lead dose, and thus bone lead content (the measured tibia lead values ranged from those seen in non-occupationally exposed modern-day humans to those seen in heavy and prolonged occupational exposure). The study was conducted *per* a protocol approved by the Wadsworth Center's Institutional Animal Care and Use Committee (IACUC). (The Wadsworth Center is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International; has an approved Animal Welfare Assurance (#A3183-01) with the United States Public Health Service; and is registered as a Class R Research Facility (#21-R-0124) with the United States Department of Agriculture.)

2.2. Preparation of the goat tibiae

Following dissection, cleaning and drying (but prior to the interlaboratory study), each tibia was marked with a cross near the mid-point of the anterior surface, using a black, water-soluble marker (the ink from which had been previously determined to have no detectable lead *via* XRF). The anterior surface was selected for XRF measurement because it was relatively flat, and the mark provided a 'target' measurement location for XRF participants. Following the conclusion of the interlaboratory study, each tibia was sectioned in a proximal-distal direction, at both 1 cm and 3 cm each side of the center of the mark (X), yielding three 2-cm sections (Fig. 1) with a diamond-coated, rotary cutting saw (Dremel, Racine, WI) that had been cleaned with dilute HCl. Each section was subsequently divided into external surface pieces of (periosteal lamellar) bone, each of approximately 2 mm thickness, and the remaining core (Haversian systems, core lamellar, and endosteal surface lamellar) bone, of approximately 1 cm thickness (Fig. 1). Surface and core samples were each further arbitrarily divided into two sub-samples (A and B) for analysis. The dry weight of the samples from each of the nine tibiae was determined after drying them at 105°C for 2 h, to remove residual water and storing them for 24 h in a vacuum desiccator over anhydrous CaSO_4 (Drierite, W.A. Hammond Drierite Company Ltd, Xenia, OH). Additional sections of the tibiae, each approximately 0.5 cm in length, were taken from the non-sectioned remainder of tibiae for determination of the ashed weight which was measured by weighing the samples before and after heating in a muffle furnace at 450°C for 48 h, to oxidize the organic matter.¹⁷ The dry and ashed weights for each tibia were measured to allow bone lead data obtained by ICP-MS to be calculated on both dry weight and ashed weight bases.

As a decomposition method, wet ashing of biological samples, *i.e.*, sample digestion in a closed vessel with concentrated mineral acids and microwave-assisted heating, is preferred

over dry ashing, *i.e.*, heating the sample in a muffle furnace, as the former is less prone to contamination errors and analyte losses due to volatilization and insoluble residues.^{18,19}

2.3. Determination of lead by acid digestion followed by ICP-MS

Following determination of the dry weight after vacuum desiccation, the samples from the nine goat tibiae were digested in approximately 5 mL of double-distilled concentrated HNO₃ (Veritas Double Distilled; GFS Chemicals, Columbus, OH) at room temperature, and then diluted with 18.2 MΩ cm deionized water.

Aliquots of bone reference materials with four different certified mass fractions for lead (New York State Reference Materials 05-01 through 05-04)²⁰ were digested in the same way, for validation, quality control and traceability purposes. Aqueous calibration standards (2% v/v HNO₃) were prepared from 1,000 mg L⁻¹ “pure lead atomic absorption standard”, a material that the manufacturer states is traceable to NIST SRM 3128 (Perkin-Elmer, Shelton, CT). An internal standard of 10 µg L⁻¹ bismuth was added to each of the samples and calibration standards. Lead was determined using an Elan DRC-plus ICP-MS (Perkin Elmer, Shelton, CT) fitted with a Meinhard nebulizer and cyclonic spray chamber (Models WE024371 and WE025221; Meinhard Glass Products, Golden, CO).

ICP-MS measurement traceability was also provided *via* analyses of Standard Reference Material (SRM) 1486 Bone Meal (National Institute of Standards and Technology, Gaithersburg, MD). Samples of this SRM were dried in accordance with the certificate instructions (including desiccation for at least 24 h), and were then digested with double-distilled concentrated HNO₃. The certified reference materials (NIST SRM 1486 and NYS RMs 05-01 through 05-04) were measured within the same ICP-MS run as the tibia samples, and an appropriate bias correction factor used to ensure an unbroken chain of calibrations. NIST SRM 1400 is a bone ash product and is certified for lead in units of µg g⁻¹ based on the ash weight, but the certified lead concentration (9.07 µg g⁻¹) is too low to be useful for validating XRF across the range of values expected.

3. Results and discussion

3.1. XRF-measured interlaboratory data for bone lead

The averages of the five replicate XRF measurements of each of the nine caprine tibiae, obtained by each of 15 measurement systems in 10 laboratories were distributed evenly across the range of lead values reported (4 µg g⁻¹ to 56 µg g⁻¹ bone mineral), without the data from any single XRF system constituting a statistical outlier.⁵ For each of the nine tibiae, however, differences between the 15 reported average values were statistically significant. Within-lab variability, assessed *via* the standard deviation (SD) for a given tibia, ranged from 0.5 µg g⁻¹ to 6.5 µg g⁻¹ bone mineral, across the nine tibiae.

Here we include results from an additional participant – the last to receive the tibiae – whose data, although returned promptly, were received too late to be included in the previously reported analysis. Table 1 shows updated descriptive statistics for the interlaboratory study data. The robust mean and robust standard uncertainty were calculated *per* Algorithm A, *i.e.*, the method described in ISO 13528 in Annex C.²¹ Briefly, individual results were ranked in increasing order:

$$(x_1, x_2, \dots, x_i, \dots, x_p)$$

Initial values of the robust average x^* and robust standard deviation s^* were calculated as:

$$x^* = \text{median } x_i \ (i=1, 2, \dots, p) \quad (1)$$

$$s^* = \text{median } |x_i - x^*| \ (i=1, 2, \dots, p) \quad (2)$$

The initial values x^* and s^* were updated by calculating:

$$\delta = 1.5 s^* \quad (3)$$

For each x_i , x_i^* was calculated where:

$$\begin{aligned} &\text{if } x_i < x^* - \delta, \ x_i^* = x^* - \delta \\ &\text{if } x_i > x^* + \delta, \ x_i^* = x^* + \delta \\ &\text{otherwise, } x_i^* = x_i \end{aligned} \quad (4)$$

New values for x^* and s^* were calculated as:

$$x^* = \sum x_i^* / p \quad (5)$$

$$s^* = 1.134 \sqrt{\sum (x_i - x^*)^2 / (p - 1)} \quad (6)$$

The robust estimates of x^* and s^* were calculated by iteration by updating the values of x^* and s^* until they converged to the third significant figure. According to ISO 13528 : 2005(E) section 5.6.2,²¹ when the assigned value is derived as a robust average x^* calculated using Algorithm A, the standard uncertainty u_x of the assigned value X is estimated as:

$$u_x = 1.25 s^* / \sqrt{p} \quad (7)$$

This approach is recommended for establishing the consensus value of an interlaboratory study where other methods of establishing an assigned or target value (*e.g.* a certified reference value) are not available (as was the case for this interlaboratory study at the time it was conducted). The standard uncertainty u_x may be expanded using a coverage factor of 2 for an approximate 95% confidence level.

3.2. ICP-MS-measured bone lead data

Table 2 summarizes the results of the ICP-MS analyses of surface and core bone samples from each of the nine test tibiae. The bone lead results given in Table 2 are reported in $\mu\text{g g}^{-1}$ dry weight units because they are based on wet ashing (digestion) of dried bone followed by lead determination by ICP-MS. The typical ratio of surface-to-core bone mass was 1 : 4. Table 2 also shows the data for 'whole' bone lead ([whole bone Pb]) that were estimated by combining the mass fraction data from surface and core components (eqn 8).

$$\text{Pb}_{\text{whole-bone}} (\mu\text{g g}^{-1}) = \frac{\text{Pb}_{\text{surf}} (\mu\text{g g}^{-1}) \times m_{\text{surf}} (\text{g}) + \text{Pb}_{\text{core}} (\mu\text{g g}^{-1}) \times m_{\text{core}} (\text{g})}{m_{\text{surf}} (\text{g}) + m_{\text{core}} (\text{g})} \quad (8)$$

$$SD1 = \sqrt{SD^2(P2a) + SD^2(P2b) + SD^2(Xa) + SD^2(Xb) + SD^2(D2a) + SD^2(D2b)} \quad (9)$$

where: Pb_{surf} is the mass fraction of lead in surface bone; m_{surf} is the mass of surface bone analyzed; Pb_{core} is the mass fraction of lead in core bone; m_{core} is the mass of core bone analyzed; and $Pb_{whole\ bone}$ is the combined mass fraction from core and surface bone.

The differences between the dry weight lead content of surface and core bone components were statistically significant, and ranged from $0.30\ \mu\text{g g}^{-1}$ (Student's *t* test, *P* value = 0.0024) to as much as $31.4\ \mu\text{g g}^{-1}$ ($P < 0.0001$), with a median of $9.5\ \mu\text{g g}^{-1}$. In Table 2, SD1 is the analytical uncertainty associated with each ICP-MS measurement, *i.e.*, the combined SD obtained by taking the square root of the sum of squares of each SD value found for each of the six 6 section/sub-samples analyzed (eqn 9).

In contrast, SD2 represents the combination of analytical uncertainty with the biological/locational variability and is simply the SD of the six mean lead values found for the 6 section/sub-sample combinations). The biological/locational variability in dry weight lead content ranged from $<0.1\ \mu\text{g g}^{-1}$ to $5.4\ \mu\text{g g}^{-1}$ ($<5\%$ to 15% relative), and from $<0.1\ \mu\text{g g}^{-1}$ to $10.7\ \mu\text{g/g}$ ($<5\%$ to 36% relative), for surface and core, respectively. Comparing the biological/locational variability to that of the analytical uncertainty (1% to 3% for all except one sample of 9% , the lowest lead content), indicates that, for all except that one surface sample, the biological/locational distribution of lead was heterogeneous, not only at the surface, but within the core too. These observations were not surprising given previous data on the heterogeneity of bone lead distribution that were obtained following the analysis of select tibiae obtained from the matching (left/right) limb for lead content prior to circulating the nine tibiae to the XRF participants (Katherine Hetter, MS Thesis, The University at Albany, 2006). The observed sample heterogeneity is also consistent with other data, wherein the microdistribution of lead within a $5\ \text{mm} \times 8\ \text{mm}$ cross-section of a caprine tibia was mapped at $80\text{-}\mu\text{m}$ -resolution *via* a benchtop monochromatic microbeam X-ray fluorescence instrument.²² In that study, lead was found to accumulate at the tibia surface unevenly and several “hot spots” were evident within the core. While the heterogeneity of lead in these caprine bones represents a limitation, the reality is that the practical application of XRF to bone lead measurement is subject to the same analyte heterogeneity when *in vivo* measurements are taken on human subjects. The circulation of these caprine bones, albeit without the overlying tissues intact, is an important step in understanding the relative contribution of all source(s) of uncertainty when XRF measurements are performed under “real world” conditions. A second XRF interlaboratory comparison that was planned as a follow up study involved circulating ground and homogenized bone lead materials (as pressed disks) that were certified for lead content earlier.²⁰ Unfortunately, only two participants returned results. We hope to repeat that interlaboratory comparison in the future as resources permit.

3.3. Matrix conversion of XRF data for comparison to ICP-MS data

We elected to convert the XRF-measured data ($\mu\text{g g}^{-1}$ bone mineral) to the mass fraction units of the ICP-MS-measured bone lead data ($\mu\text{g g}^{-1}$ dry weight), rather than *vice versa*, because all but one of the few available reference materials are certified for lead content in mass fraction units of $\mu\text{g g}^{-1}$ dry weight. We elected to not standardize on the native units of SRM 1400 because it's of limited use.

Conversion of XRF-measured values can be performed in more than one way; the bone mineral data reported by the XRF labs could be: (a) divided by the matrix (coherent) conversion factor reported by the lab, to return them to plaster-based mass fraction units,

and then multiplied by a plaster-to-dry-weight-bone conversion factor; (b) multiplied by the average ash fraction from the nine tibiae; or (c) multiplied by the ash fraction of the particular tibia measured. All three methods were performed and yielded almost identical numbers and no discernable difference. Method (b) was adopted because method (a) is subject to rounding in the reported conversion factor values, and because the ash fractions of the tibiae showed smaller variability (4% RSD) compared to the biological uncertainty, which can be more than 10% RSD. Finally, we recognize the potential for difference in mineral content between the bone segments analyzed by XRF and then wet ashed, *versus* the segments that were dry ashed, but consider any such effect to be likely small because there was no discernable difference in the conclusions drawn between using the average ash fraction and using the tibia-specific ash fraction.

The experimental determination of the fractional weight loss during dry ashing (*i.e.*, the quotient of the weight of the bone mineral and the dry weight of the bone) was 0.676 (SD 0.027 (4%), standard error of the mean (SEM) 0.009 (1.3%)). This experimentally determined value is similar to the value of 0.638 than can be inferred from the data of Woodard and White, who reported ash and protein masses, along with weight not accounted for, for up to 20 determinations from as many as nine cortical (humerus, femur or tibia) bones from human subjects aged 5 years to 72 years (and which Woodard and White state to be a function of age). It is also similar to the value of 0.605 for tibia reported by Wittmers *et al.*¹⁷ XRF practitioners might also be interested to note that the observed ash fraction of 0.676 is also close to the inverse of the typical coherent conversion factor ($1/1.462 = 0.684$) – the factor by which to scale $\mu\text{g g}^{-1}$ bone mineral back to $\mu\text{g g}^{-1}$ plaster – but it should be noted that the coherent conversion factor is independent of the amount of non-mineral (*i.e.*, organic) matter in the bone, whereas the ICP-MS measurements are not because the organic components of bone contribute to the weighed mass. This dependence of “dry weight” measurements on the organic component of bone is another source of uncertainty that must be considered given that XRF (specifically, the coherent conversion factor) is negligibly influenced by interactions with the organic matter.

3.4. Comparison of XRF and ICP-MS data

XRF-measured lead data, converted-to-dry-weight, and ICP-MS-measured lead data (surface, core and whole bone) are given in Table 3 and the differences between the XRF- and ICP-MS-measured data are represented graphically in Fig. 2 as a difference plot. A more detailed description of differences among the XRF systems, and analysis of the interlaboratory XRF data have been reported elsewhere.⁵

Our comparisons between XRF and ICP-MS as described here are more detailed (and somewhat more complicated) by the separation of surface and core bone sections. If the analyzed mass of bone had been rendered homogenized (by grinding, for example) prior to ICP-MS analysis, rather than having been separated into physiologically distinct surface and core bone components, ICP-MS analysis would have yielded a single estimate of mass fraction (likely close to the whole bone value reported here) and a smaller SD (likely close to the SD1 values) but such an approach would have failed to observe the substantial and important variability in the lead content in bone that we report here. This biological variability, and the difference between surface and core are important, both *per se* and for XRF, and so we preferred to not lose this information by homogenizing the samples.

The validity of the comparisons between XRF and (each of) surface, core and whole bone lead measured *via* ICP-MS is influenced by the energy and consequent ability of the fluorescing photons to penetrate into bone (and through soft tissue in the *in vivo* situation), and the fluoresced (K-shell) lead x rays to escape from bone (and soft tissue). For ¹⁰⁹Cd-based XRF, these energies result in sampling of both surface and core bone; sampling

surface bone with more sensitivity than that of core bone; and, consequently, neither surface, core, nor whole bone being exactly what XRF samples, *viz.* surface-bone-weighted-whole bone. Some ‘under-sampling’ of the core would thus be expected (and was observed in our earlier study of human cadaver intact legs (Todd *et al.* 2002)). Indeed, in an experiment using a lead wire positioned at various points within a water ‘phantom’, Somerville *et al.* observed a reduction of 58% in relative detection efficiency at 2 cm distance, as compared to the efficiency at the surface directly opposite the XRF source, and a decline >90% at 3 cm.⁸ Given that greater attenuation of x rays will occur in bone compared to water, a steeper decline in relative detection efficiency is expected. The 6 cm length of tibia sampled in the current study is also likely not exactly the same extent of bone sampled by XRF, but is probably not dissimilar, and the thickness of the goat tibiae (~2 cm) is less than the mean free path of a lead K α_1 x ray (~22 mm). This indicated (K)XRF samples the entire depth of the goat bone, and even beyond, in these experiments. Given that the exact mass/volume of bone assessed by each of the XRF measurement systems cannot be definitively known, the heterogeneity of the lead distribution must be considered as a potential source of disagreement in these results. Nevertheless, because XRF is assumed to sample a mass/volume of bone similar to the mass/volume digested for ICP-MS, the mean XRF result should end up close to the ICP-MS mean. With these caveats in mind, the comparisons between XRF and ICP-MS can be considered.

Comparing XRF to whole bone ICP-MS, these data show, qualitatively speaking, good agreement over the range of tibia lead mass fractions assessed ($2 \mu\text{g g}^{-1}$ to $36 \mu\text{g g}^{-1}$ dry weight, or, more usefully for XRF practitioners, approximately $4 \mu\text{g g}^{-1}$ to $55 \mu\text{g g}^{-1}$ bone mineral). XRF exceeded whole bone ICP-MS for tibiae 82-17 (by $1.6 \mu\text{g g}^{-1}$ dry weight, 4.5%) and 89-4 (by $3.4 \mu\text{g g}^{-1}$ bone mineral, 11.8%), the two highest lead values, but neither bias achieved statistical significance ($P = 0.245$ and 0.153 , respectively).

With the exception of the tibia (96-1) with the lowest lead content, the ICP-MS surface lead data were higher than their corresponding mean XRF results. Thus, in this study, XRF underestimates surface lead content but, given that XRF samples more than surface bone, this underestimation does not represent a deficiency in the XRF method. That said, the difference between surface lead and XRF-measured lead was statistically significant for all but the two tibiae with the lowest lead content. In our earlier XRF study of human cadaver intact legs,¹² there was no statistically significant difference between XRF and surface lead, but that study was performed with the samples intact (*i.e.* with skin and overlying tissue still in place) and assayed bones of a more confined range in lead content ($3 \mu\text{g g}^{-1}$ to $17 \mu\text{g g}^{-1}$ dry weight in the core, and $9 \mu\text{g g}^{-1}$ to $28 \mu\text{g g}^{-1}$ dry weight in the surface *via* AAS) than those assessed in this study, indicating lower exposure of the humans (compared to the goats), which would result in larger differences between XRF and surface lead at higher lead mass fractions, if surface enrichment is a proportional phenomenon.

Similarly, XRF does not sample core bone alone, although the greater mass of core than surface bone sampled by ICP-MS suggests, *a priori*, that XRF data would agree better with core than surface. This is borne out in the comparison of the data; XRF exceeded core bone lead for all except one of the tibiae, but those differences achieved statistical significance for only two of the nine tibiae (four other differences were of borderline significance, $0.05 < P < 0.1$). Statistical significance notwithstanding, the differences between XRF and core lead ranged from $+0.5 \mu\text{g g}^{-1}$ to $-9.8 \mu\text{g g}^{-1}$ dry weight bone (median $2.1 \mu\text{g g}^{-1}$).

Fig. 3 and 4 show the difference between XRF and whole bone lead for each of the individual XRF measurement systems. The grey bars represent the combined uncertainty arising from both analytical sources and the heterogeneous locational/biological distribution of lead within the measured mass/volume of each individual tibia. Eight of the systems did

not yield a significant bias for any of the nine tibiae; one system was biased for one of the tibiae; two systems were biased for two tibiae; one system was biased for four tibiae; two systems (813-1 and 804-2) were biased for five tibiae and one system (801-1) was biased for six of the nine tibiae. Average bias for the systems (under those particular operating conditions) that were biased for the majority of samples were $5.1 \mu\text{g g}^{-1}$ (30.7%), $-2.6 \mu\text{g g}^{-1}$ (-15.7%) and $4.6 \mu\text{g g}^{-1}$ (27.7%) dry weight bone ($7.6 \mu\text{g g}^{-1}$, $-3.8 \mu\text{g g}^{-1}$ and $6.9 \mu\text{g g}^{-1}$ bone mineral) for systems 813-1, 804-2 and 801-1, respectively. All participants now have the ICP-MS data, allowing any corrective actions deemed necessary to be implemented.

4. Conclusions

The 16 XRF systems that were used in the interlaboratory comparison study produced results for lead in bone that are 'fit for purpose', *i.e.* *in vivo* measurements of bone lead. It is recognized that there is an inherent compromise between the high accuracy, high precision data that is typically associated with inorganic mass spectrometry and the kind of data that is achievable based on *in vivo* KXRF, coupled with biological variability – which we cannot alter. The range between positively biased and negatively biased systems was such that researchers must use caution, or correct, when comparing studies conducted with different systems. There is a clear need for common reference materials available to all XRF laboratories, to provide comparison across systems and internal quality control.

Acknowledgments

This project was supported in part by grant R01 ES12424 from the National Institute of Environmental Health Sciences (NIEHS), a division of the National Institutes of Health (NIH), USA. The contents of this paper are, however, solely the responsibility of its authors and do not represent official views of the NIEHS or NIH. The authors gratefully recognize the contribution of Dr F.S. Blaisdell, Facility Veterinarian, and the staff of the Trace Elements Laboratory at the Wadsworth Center involved in the animal care, lead dosing and collection and archiving of goat materials over many years.

References

1. Hu H, Shih RA, Rothenberg SJ, Schwartz BS. Environ. Health Perspect. 2006; 115:455–462. [PubMed: 17431499]
2. Navas-Acien A, Guallar E, Silbergeld EK, Rothenberg SJ. Environ. Health Perspect. 2006; 115:472–482. [PubMed: 17431501]
3. Shih RA, Hu H, Weisskopf M, Schwartz BS. Environ. Health Perspect. 2006; 115:483–492. [PubMed: 17431502]
4. Schwartz BS, Hu H. Environ. Health Perspect. 2006; 115:451–454. [PubMed: 17431498]
5. Parsons PJ, Bellis DJ, Hetter KM, Geraghty C, Berglund NA, Ginde NR, Mata P, Todd AC. X-Ray Spectrom. 2008; 37:76–83.
6. Woodard HQ. Health Phys. 1962; 8:513–517. [PubMed: 14001769]
7. Somervaille LJ, Chettle DR, Scott MC, Tennant DR, McKiernan MJ, Skilbeck A, Trethowan WN. Br. J. Ind. Med. 1988; 45:174–181. [PubMed: 3348993]
8. Somervaille LJ, Chettle DR, Scott MC. Phys. Med. Biol. 1985; 30:929–943. [PubMed: 4048276]
9. Todd AC. Phys. Med. Biol. 2000; 45:1953–1963. [PubMed: 10943931]
10. Gordon CL, Webber CE, Chettle DR. Environ. Health Perspect. 1994; 102:690–694. [PubMed: 7895710]
11. Todd, AC.; McNeill, FE. Human Body Composition Studies. Ellis, KJ.; Eastman, JD., editors. New York: Plenum Press; 1993. p. 299–302.
12. Todd AC, Parsons PJ, Carroll S, Geraghty C, Khan FA, Tang S, Moshier EL. Phys. Med. Biol. 2002; 47:673–687. [PubMed: 11900198]

13. Aro A, Amarasiriwardena C, Lee ML, Kim R, Hu H. *Med. Phys.* 2000; 27:119–123. [PubMed: 10659745]
14. Hu H, Milder FL, Burger DE. *Arch. Environ. Health.* 1990; 45:335–341. [PubMed: 2270952]
15. Somervaille LJ, Chettle DR, Scott MC, Aufderheide AC, Wallgren JE, Wittmers L Jr, Rapp GR Jr. *Phys. Med. Biol.* 1986; 31:1267–1274. [PubMed: 3786412]
16. Parsons PJ, Reilly AA, Esernio-Jenssen D, Werk LN, Mofenson HC, Stanton NV, Matte TD. *Clin. Chem.* 2001; 47:322–330. [PubMed: 11159782]
17. Wittmers L Jr, Wallgren JE, Alich A, Aufderheide AC, Rapp GR Jr. *Arch. Environ. Health.* 1988; 43:381–391. [PubMed: 3196073]
18. Mader P, Szakova J, Miholova D. *Analisis.* 1998; 26:121–129.
19. Novozamsky I, Vanderlee HJ, Houba VJG. *Microchim. Acta.* 1995; 119:183–189.
20. Bellis DJ, Hetter KM, Verostek MF, Parsons PJ. *J. Anal. At. Spectrom.* 2008; 23:298–308.
21. ISO 13528. Geneva, Switzerland: International Standards Organisation; 2005. Statistical Methods for use in Proficiency Testing by Interlaboratory Comparisons.
22. Bellis DJ, Li D, Chen Z, Gibson WM, Parsons PJ. *J. Anal. At. Spectrom.* 2009; 24:622–626.

Fig. 1. Schematic representation of the sampling analytical protocols used to obtain bone ash data and bone lead data based on wet ashing of dry bone. The center of the tibiae is denoted as "X".

Fig. 2.

Difference plot comparing the *in vitro* surface, core, and whole (*i.e.*, surface and core, combined) lead data ([Pb]), determined *via* ICP-MS to the robust mean lead value from the XRF interlaboratory study. [Pb]; lead mass fraction.

Fig. 3.
Part 1: Systems 801-1, 801-2, 803-1, 803-2, 804-1, 804-2, 806-1, 806-2. Difference plots comparing the XRF-measured lead values, ([Pb]), to the mean value determined *via* ICP-MS. The error bars indicate the standard deviations of five XRF replicate measurements. The grey bars represent the uncertainty arising from locational-biological inhomogeneity and analytical reproducibility (ICP-MS SD2).

Fig. 4. Part 2: Systems 802-1, 808-2, 812-1, 812-2, 810-1, 811-1, 813-1, 815-1. Difference plots comparing the XRF-measured lead values, ([Pb]), to the mean value determined *via* ICP-MS. The error bars indicate the standard deviations of five XRF replicate measurements. The grey bars represent the uncertainty arising from locational-biological inhomogeneity and analytical reproducibility (ICP-MS SD2).

Table 1

Lead mass fractions ($\mu\text{g g}^{-1}$ bone mineral) for nine caprine tibiae, measured by 16 XRF systems in 11 laboratories in an interlaboratory study. Each value is the mean from 16 systems, for each of which there were five consecutive replicate measurements made of each tibia. The robust mean and standard uncertainty are calculated *per* ISO 13528²¹

Tibia	Mean	Minimum	Maximum	SD	Robust Mean	Standard Uncertainty
82-17	55.3	47.6	63.5	4.7	55.3	1.6
86-6	28.9	19.1	37.1	4.9	28.9	1.5
89-4	47.5	27.6	60.0	8.1	47.8	2.3
89-14	24.1	19.3	31.2	4.0	23.8	1.2
89-15	23.0	12.9	31.4	4.8	22.6	1.2
93-2	12.5	2.0	21.6	4.9	12.1	1.0
93-6	21.9	16.6	29.9	3.9	21.6	1.2
95-4	11.9	7.4	19.3	3.8	11.3	0.9
96-1	4.0	-2.1	9.5	3.1	4.0	1.0

Table 2

Lead mass fractions (in $\mu\text{g g}^{-1}$ dry weight bone) for nine caprine tibiae, measured *via* ICP-MS. Each value is the mean of six measurements (surface and core sub-sections for each of three proximal-distal sections). Uncertainty is expressed as (a) SD1, the combined SD of the six ICP-MS measurements (see eqn (9) in text), and (b) SD2, the SD of the values measured in each of the six different regions of the tibia

Tibia	Surface			Core			Whole		
	Mean	SD1	SD2	Mean	SD1	SD2	Mean	SD1	SD2
82-17	61.2	1.04	6.4	29.8	0.22	10.9	35.8	0.7	10.2
86-6	26.2	0.85	3.2	16.4	0.10	3.6	18.7	0.6	3.5
89-4	48.5	0.97	5.3	22.5	0.16	4.5	28.9	0.7	4.7
89-14	22.9	0.66	2.1	13.9	0.12	1.6	15.9	0.5	1.7
89-15	23.0	0.57	2.1	13.5	0.18	1.1	15.7	0.5	1.4
93-2	10.0	0.29	1.8	8.7	0.08	1.1	9.0	0.2	1.3
93-6	22.4	0.62	3.9	12.5	0.11	2.5	15.2	0.5	3.0
95-4	12.9	0.41	1.6	6.4	0.04	2.0	8.0	0.3	1.9
96-1	2.0	0.18	0.2	1.7	0.03	0.1	1.8	0.1	0.1

Table 3

ICP-MS- and XRF-measured lead mass fractions and the differences between them (all in units of $\mu\text{g g}^{-1}$ dry weight bone) for nine caprine tibiae measured, for XRF, by 16 XRF systems in 11 laboratories^a

Tibia	ICP-MS-measured bone [Pb]				XRF-measured bone [Pb]					
	Mean	SD1	SD2	Mean	Min.	Max.	SD	Robust Mean	Standard Uncertainty	XRF - ICP-MS ^b
82-17	35.8	0.7	10.2	37.4	32.2	42.9	3.2	37.4	1.1	1.6
86-6	18.7	0.6	3.5	19.5	12.9	25.1	3.3	19.5	1.0	0.8
89-4	15.9	0.5	1.7	32.1	18.7	40.6	5.5	32.3	1.6	16.4
89-14	15.7	0.5	1.4	16.3	13.0	21.1	2.7	16.1	0.8	0.4
89-15	28.9	0.7	4.7	15.5	8.7	21.2	3.2	15.3	0.8	-13.6
93-2	9.0	0.2	1.3	8.5	1.4	14.6	3.3	8.2	0.7	-0.8
93-6	15.2	0.5	3.0	14.8	11.2	20.2	2.6	14.6	0.8	-0.6
95-4	8.0	0.3	1.9	8.0	5.0	13.0	2.6	7.6	0.6	-0.4
96-1	1.8	0.1	0.1	2.7	-1.4	6.4	2.1	2.7	0.7	0.9

^a [Pb]: lead mass fraction; SD1: analytical uncertainty; SD2: analytical uncertainty + biological/location variability.

^b the difference between the robust mean of the XRF-measured lead values and ICP-MS-measured bone lead values.