

Age estimation of immature human skeletal remains from the diaphyseal length of the long bones in the postnatal period

Hugo F. V. Cardoso · Joana Abrantes ·
Louise T. Humphrey

Received: 1 August 2013 / Accepted: 30 September 2013 / Published online: 15 October 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract Age at death in immature human skeletal remains has been estimated from the diaphyseal length of the long bones, but few studies have actually been designed specifically for the purpose of age estimation and those which have, show important caveats. This study uses regression and classical calibration to model the relationship between age and diaphyseal length of the six long bones, in a sample of 184 known sex and age individuals (72 females and 112 males), younger than 13 years of age, selected from Portuguese and English skeletal collections. Age estimation models based on classical calibration were obtained for each of the six long bones, and separately for each sex and for the sexes combined, and also for the entire sample and when it is subdivided into two subsamples at the age of 2 years. Comparisons between inverse and classical calibration show there is a systematic bias in age estimations obtained from inverse calibration. In the classical calibration models, the length of the femur provides the most accurate estimates of age. Age estimates are more accurate for the male subsample and for individuals under the age of 2 years. These results and a test of previously published methods caution against inverse calibration as a technique for developing age estimation methods even from the immature skeleton. Age estimation methods developed

using cemetery collections of identified human skeletons should not be uncritically applied to present-day populations from the same region since many populations have experienced dramatic secular trends in growth and adult height over the last century.

Keywords Age at death · Growth · Regression · Classical calibration · Secular trend

Introduction

Age estimation is not only an essential first step in the identification of unknown human skeletal remains in a forensic setting but also in the analysis of archaeological samples of human bones. Age estimation is most accurate for immature skeletons, and there are three main approaches to age estimation: dental development, skeletal growth, and skeletal maturation [1]. The most reliable methods for ageing are based on tooth formation schedules, [1–4] but there are circumstances in which they cannot be used. The linear growth of the skeleton, particularly the diaphyseal growth in length of the long bones is generally considered to be a good alternative to dental mineralization for age estimation in prepubertal children [5–8].

Although age of immature human skeletal remains has been recurrently estimated from long bone diaphyseal lengths, using a variety of methods and/or approaches [9–29], most data available are unsuitable for age prediction. Several publications include tables of descriptive statistics for bone lengths by age, but these studies were designed to determine the mean long bone length for a given age when assessing growth status in living children. Radiographic data for long bone length by age is the source of this information, and detailed tables have been published by Maresh [9–11], Ghantus

H. F. V. Cardoso (✉)

Department of Archaeology, Simon Fraser University, Burnaby,
British Columbia V5A 1S6, Canada
e-mail: hcardoso@sfu.ca

J. Abrantes

Departamento de Medicina Legal e Ciências Forenses, Faculdade de
Medicina, Universidade do Porto, Porto 4200-319, Portugal

L. T. Humphrey

Human Origins Group, Department of Palaeontology, The Natural
History Museum, London SW7 5BD, UK

[12], Anderson and coworkers [13], and Gindhart [14]. A few tables of descriptive statistics based on actual measurements of dry bones have been constructed as an aid for age estimation, including those published by Johnson [15], Walker [16] Stloukal and Hanáková [17], Sundick [18], and Hoppa [19], but the samples are only of estimated age as they are archaeological in origin. Fazekas and Kosa [20] also provide tables of descriptive statistics for dry bone material, but in this case for the fetal period only.

More recent studies have been concerned with exploring and modeling the relationship of long bone length with age, such as the studies carried out by Smith and Buschang [30, 31]. These studies are based on Maresh's [11] original data and have been incorrectly described as age estimation methods [7] as the models were designed to estimate the average long bone length for a given age, and not the age from a given long bone length. Predictive models with the specific purpose of age estimation from length of long bones were initially developed by Stewart [21] and Hoffman [22], but here age and the respective confidence intervals have to be extrapolated from graphed data of diaphyseal lengths. More recently, Facchini and Veschi [23], Rissech and coworkers [24–26], Danforth and coworkers [27], Boccone and coworkers [28], and Primeau and coworkers [29] have all used regression analysis for age estimation purposes. There are some important caveats with all these methods. The regression formulae provided by Danforth and coworkers [27], Boccone and coworkers [28], and Primeau and coworkers [29] were based on an archaeological sample of unknown age, where ages were estimated. By contrast, Facchini and Veschi [23] and Rissech and coworkers [24–26] used known sex and age skeletal collections to derive their formulae. Some of these formulae have been published without reporting error estimates and, hence, cannot be used to obtain a confidence interval for the estimated age. This applies to all formulae derived by Facchini and Veschi [23] and the femur formulae obtained by Rissech and coworkers [24]. Consequently, at present only the humerus and tibia formulae published by Rissech and coworkers [25, 26] can be considered suitable for age estimation from the diaphyseal length of the long bones, as they allow the estimation of error. Conversely, as important as they are, these equations were developed using conventional least squares regression to produce age estimates and are likely to introduce significant biases [32].

In age estimation, the common model used is least squares regression and inverse calibration. Age is the dependent variable and the long bone length is the independent variable, but it is age that is unknown and needs to be estimated from long bone length. When inverse calibration is used to estimate age, age (the dependent variable or x) is actually regressed on long bone length (the independent variable or y), rather than the reverse. Consequently, random errors in this inverse calibration are assumed entirely in the y direction, when in fact they

are in the x direction. Considering that least squares was designed to minimize errors in the y direction, inverse calibration where age is treated as the independent variable (y) and long bone length is treated as the dependent variable (x), results in a systematic bias [32]. Konigsberg and coworkers [33] and Lucy and Pollard [34] have recommended classical calibration as a more suitable statistical technique for making age estimates from skeletal indicators. In classical calibration, the variable for which estimates are to be made is always x (age), not y (long bone length) as in inverse calibration. A regression y (long bone length) on x (age) is performed as usual (not age regressed on long bone length as in inverse calibration). However, this produces an equation for long bone length (y) in terms of age (x), so to estimate age we must invert the relationship. In this case, least squares regression adjusts the regression equation in the correct y direction. The major drawbacks in classical calibration are the difficulty in calculating the uncertainty for any point about the calibration line and a reduction in the efficiency of estimates [32, 33], that is, the variability will be larger for classical calibration than for inverse calibration. Although this problem has been described exclusively for adult age estimation, it is likely that it will affect age predictions from immature skeletal remains as well, albeit at a smaller level. According to Aykroyd and coworkers [32] the greater the correlation between age and the skeletal indicator, less systematic bias will be in the inverse calibration model. Given the low correlation between age and skeletal indicators of age in adults, it is no surprise that the conventional use of least squares regression is of great concern. Conversely, one might assume that the bias associated with inverse calibration is negligible when applied to the immature skeleton given the very high correlation between age and long bone length. However, considering that this correlation is not perfect, inverse calibration may not necessarily translate into a bias free regression model for age estimation. In fact, we would still expect at least some bias.

This study addresses the issue of modeling age and long bone length for age estimation purposes, using the most appropriate statistical tools that will not result in methodological biases. Consequently, a sample comprised of immature individuals from Western European documented skeletal collections (Portugal and England), were selected to develop age estimation formulae from the diaphyseal length of the humerus, radius, ulna, femur, tibia, and fibula in prepubertal children (<12 years of age) using regression and classical calibration [35]. This study compares the differential performance of inverse calibration versus classical calibration formulae. An additional goal is to determine which bones are most accurate in estimating age and whether there are sex and age differences in the accuracy of these formulae. One last goal is to test previously published equations for age estimation from diaphyseal lengths of the long bones.

Materials and methods

Sample

For this study, data were collected from two samples of child skeletons of known sex and age: the Portuguese sample, which includes the Lisbon collection children, and the English sample, where the Spitalfields and St. Bride's children were combined. The Lisbon collection is a large series of identified human skeletons housed at the Natural Museum of Natural History and Science, in Lisbon, Portugal [36]. This collection is composed of over 1,500 skeletons but detailed biographic information is available only for a fraction of the individuals. These fully identified skeletal remains are of Portuguese nationals' who were born between 1805 and 1972 and died between 1880 and 1975 in and around the city of Lisbon [34]. The Spitalfields [37] and St. Bride's [38] collections are, respectively, curated in the Natural History Museum and the crypt at St. Bride's Church in Fleet Street, London, UK. The collections consist mostly of Londoners who were born and died between 1729 and 1859. The Spitalfields collection includes 968 individuals and the St. Bride's collection includes 237 individuals.

As this study is based on diaphyseal length, the sample was restricted to skeletons of prepubertal age and showing no evidence of fusion of any of the long bone epiphyses. Consequently, only individuals under the age of 13 were selected. Considering the absence of significant fetal material, the study sample is also truncated inferiorly at birth. Specimens with obvious skeletal malformations were not included. In total, the study sample is comprised of 184 individuals (72 females and 112 males) with ages ranging from birth to 12 years and is of known sex. Overall, the sample is comprised of children who were born and died between approximately 250 and 50 years ago in Western Europe (Portugal and England). Table 1 describes the size and composition of the sample.

Methods

The maximum diaphyseal length of the six long bones of the limbs—humerus, radius, ulna, femur, tibia, and fibula—was measured using an osteometric board or a sliding caliper, in the case of infant skeletons, and recorded to the next whole millimeter. Measurements were obtained from the left side as the maximum length of the diaphysis, parallel to the long axis [39]. When bones from the left side were missing or damaged, the bones from right side were measured instead. Intra- and interobserver measurement errors were estimated by re-measuring a random subsample of 20 individuals, and calculating the relative technical error of measurement (%TEM) and the coefficient of reliability (R) [40] for each bone.

In all subsequent analyses, the sample was separated by sex and divided into two age groups. One subsample included all

individuals younger than 2 years of age (<2 years) and the other subsample all individuals 2 years of age and older (≥ 2 years). The division of the sample at the age of 2 years reflects biological realities of the growth process. After birth, linear growth is very fast and around the age of 2 years it slows down up to puberty [41]. Given this sharp decrease in growth velocity, the growth curve may not properly be modeled by simple linear regression between birth and 12 years of age. Consequently, this separation allowed the sample to be modeled separately for growth before and after the age of 2 years, using linear regression (Fig. 1). As the simple linear regression and calibration models are very effective and easier to calculate and use, they were preferred over nonlinear models.

Normality and homoscedasticity of the samples were tested at the start of the statistical analysis. The samples from the two series (Portuguese and English) were then compared using an ANCOVA, to determine whether there were significant size and sex differences. Subsequently, age estimation formulae were calculated using classical and inverse calibration models [35], for each long bone length, separately for the total sample, and for the subsample which included individuals younger than 2 (<2) and 2 years of age and older (≥ 2), and by sex. For the inverse calibration formulae, the standard error of the estimate (SEE) and the coefficient of determination (R^2) were calculated. The SEE cannot be obtained in classical calibration models, and in this case, the mean standard error (MSE) was calculated instead. This statistic was obtained by calculating the standard error for each individual observation as suggested by Lucy [35] and then averaging the observations to obtain a MSE for the entire sample. This is an approximation of the SEE calculated from inverse calibration [35]. Classical calibration models produce an equation for long bone length (y) in terms of age (x), so to estimate age the relationship has to be inverted and the formulae obtained is solved for age. For each classical calibration formulae, the mean long bone length and respective sample size (N), standard deviation (SD), and range (minimum and maximum), were also provided. The minimum and maximum provide the valid range of values from which age can be estimated using each of the models.

Accuracy and bias of the classical and inverse calibration models was tested on the study sample. For each long bone model, the estimated age obtained was compared with the known chronological age and both the mean residuals (MR) and mean of the absolute value of the residuals (MAR) were calculated, as an estimate of bias and accuracy respectively. Additionally, the percentage of individuals whose chronological age falls within the 95 % confidence interval (95 % CI) of the estimated age (using the MSE for the classical calibration models and the SEE for the inverse calibration models) was also calculated for each long bone length. Confidence intervals in the classical calibration models are calculated by multiplying the standard error by the appropriate value from

Table 1 Size and composition of the sample by age, sex, and collection

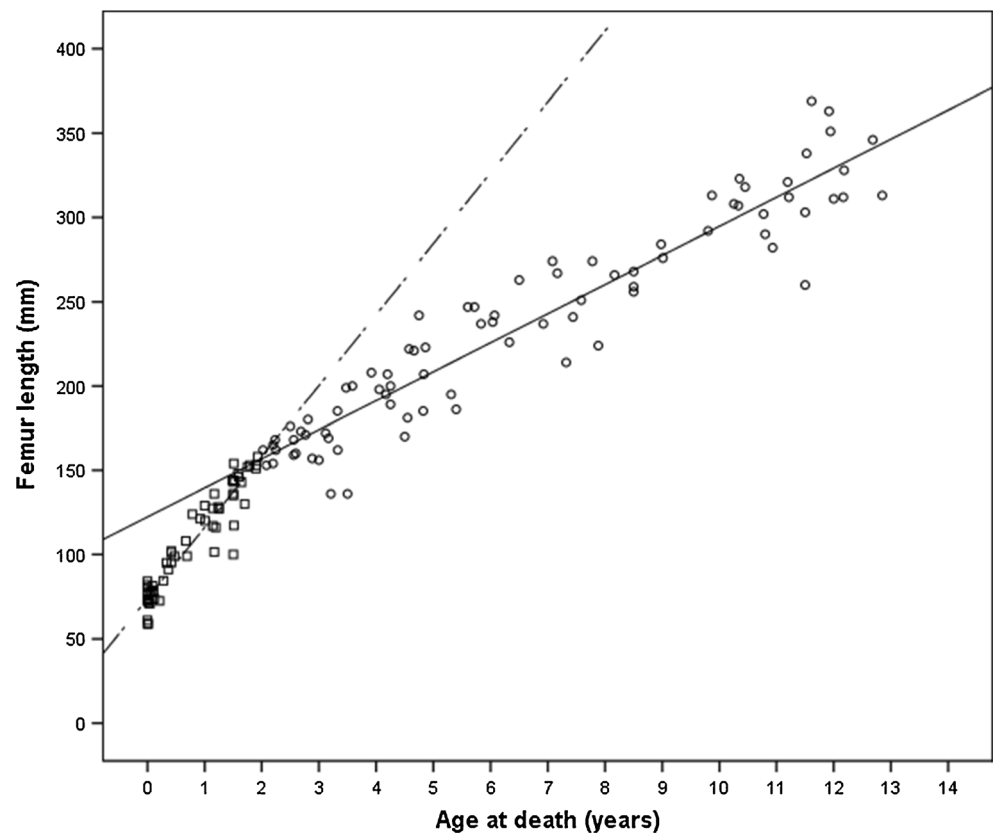
Age (years)	Lisbon collection		Spitalfields collection		St. Bride's collection		Total	
	Male	Female	Male	Female	Male	Female	Male	Female
0.0–0.09	10	11	15	11	1	0	26	22
1.0–1.9	15	7	9	6	2	0	26	13
2.0–2.9	8	12	7	3	0	0	15	5
3.0–3.9	5	3	3	3	1	1	9	7
4.0–4.9	4	5	3	0	1	0	8	5
5.0–5.9	2	2	1	0	0	0	3	2
6.0–6.9	1	2	0	0	2	0	3	2
7.0–7.9	4	1	1	0	1	0	6	1
8.0–8.9	2	2	0	0	1	0	3	2
9.0–9.9	2	2	0	0	1	0	3	2
10.0–10.9	2	4	1	0	0	0	3	4
11.0–11.9	5	4	0	0	0	0	5	4
12.0–12.9	2	2	0	1	0	0	2	3
Total	62	47	40	24	10	1	112	72

the t distribution for $n-2$ degrees of freedom [35], which is approximately 2 for all models.

Finally, previously published models for age estimation from long bone lengths were tested on the study sample, specifically, the equations provided by Rissech and coworkers [24–26] for age estimation from the femur, the tibia and the

humerus, and the equations provided by Facchini and Veschi [23] for age estimation from the humerus, the radius, the ulna, the femur, the tibia and the fibula. The accuracy of these models was tested by calculating MR and MAR. One sample t tests were used to test if MR are significantly different from zero. The percentage of individuals whose chronological age

Fig. 1 Scatterplot illustrating a classical calibration model where long bone length (femur) is regressed on age, with two separate regression lines adjusted to the data using least squares. One line is adjusted to the data of children under 2 years of age, and the other line to the data of children 2 years of age and older, when the sexes are combined. Note differences in the slope (growth velocity) and in the dispersion of data points about the regression line



falls within the 95 % CI of the estimated age was only calculated for the tibia and humerus equations provided by Rissech and coworkers [25, 26] as Facchini and Veschi [23] do not provide the SEE for their formulae and neither does Rissech and coworkers [24] for the femur.

Results

Table 2 shows the results of the intra- and interobserver measurement tests. For intra-observer error, all variables had %TEM values under 0.77 and R values equal to 1.00. The inter-observer error test results are very similar with all variables showing %TEM values under 0.72 and R values also equal to 1.00.

The ANCOVA test results show that the samples from the two series (Portuguese and English) differ significantly in long bone length by age (Table 3). However, this holds only for the subsample of individuals 2 years of age and older. The two series do not differ in size when children under the age of 2 are considered. Despite the differences between samples, they were combined into one for all subsequent analyses. No significant differences in long bone length were found between the sexes for the total sample and for the subsamples which included younger and older than 2 years of age, with the exception of the radius and ulna in the later subsample (Table 3). Although females and males tend not to differ in long bone length by age, the sexes were treated separately in the analysis, as well as combined. The ANCOVA results also showed that there is no significant interaction between the effects of sex and series.

The classical calibration models for each of the long bone lengths are shown in Tables 4, 5, and 6 for the total sample, and for the subsamples which include only individuals younger than 2 years of age (<2 years) and individuals 2 years of age and older (≥ 2 years), respectively. Each model includes the N , the regression formula solved for age (inverted), the MSE, the R^2 , the mean long bone length (M) and its respective SD and range (minimum–maximum). Given the differential preservation, sample sizes vary. Femoral diaphysis length consistently provides the best estimates of age in the total sample (MSE=1.06 years when the sexes are combined, MSE=0.92 in males and MSE=1.21 in females) and in the

subsamples which include only individuals younger than 2 years of age (MSE=0.23 years when the sexes are combined, MSE=0.23 in males and MSE=0.25 in females) and individuals 2 years of age and older (MSE=1.16 years when the sexes are combined, MSE=0.97 in males and MSE=1.40 in females). The next best bone length is the tibia, but not consistently, as the humerus and the fibula show lower MSEs in some subsamples. In fact, the tibia is the worst performing bone length in the subsample of under 2-year olds. Overall, the ulna shows the largest amount of error. Females show generally larger MSEs with the exception of children under the age of 2. The error associated with these formulae also increases with age, as the subsample of individuals younger than 2 years show a MSE of about 0.26 years, whereas in the subsample of individuals 2 years of age and older the MSE is around 1.40 (five times as much).

When testing the accuracy of classical (Table 7) and inverse calibration (Table 8) models in the study sample, classical calibration formulae show no mean difference between estimated and chronological age (MR=0.00), whereas inverse calibration formulae show consistent differences between estimated and chronological age (MR ranges between -0.50 and 0.99). In the inverse calibration model, MR seem smaller for the humerus (and overall for the upper limb bones) and largest for the femur (and overall for the lower limb bones), but not consistently. MR also tend to be smaller and negative in younger children (<2 years) and larger and positive in older children (≥ 2 years). Most of the MR are significantly different from zero ($p < 0.05$), with a few exceptions. Comparing both models (classical and inverse calibration) in terms of the percentage of individuals whose chronological age is within the 95 % CI the results vary between 90.3 and 100 % for the classical calibration model and between 57.9 and 100 % for the inverse model. This percentage tends to be slightly greater in the classical calibration models, with some exceptions, such as the femur, tibia, and fibula in the older males (≥ 2 years). In relation to MAR, the classical and inverse calibration models show similar results, but the classical calibration model has overall slightly smaller MARs. In the total sex-combined samples, MAR is around 0.95 (values vary between 0.16 and 1.38) for the classical calibration model and 0.96 (values vary between 0.17 and 1.42) for the inverse calibration model. This difference is more noticeable in the subsample which

Table 2 Intra- and interobserver measurement error test results for length of each long bone, estimated from the relative technical error of measurement (%TEM) and the coefficient of reliability (R)

	Humerus		Radius		Ulna		Femur		Tibia		Fibula	
	TEM%	R	TEM%	R	TEM%	R	TEM%	R	TEM%	R	TEM%	R
Intraobserver error	0.28	1.00	0.59	1.00	0.29	1.00	0.42	1.00	0.32	1.00	0.77	1.00
Interobserver error	0.34	1.00	0.43	1.00	0.41	1.00	0.72	1.00	0.60	1.00	0.66	1.00

Table 3 ANCOVA results for comparisons of linear regression models between the samples from the two series (Portuguese and English) and the sexes

Effects	Humerus		Radius		Ulna		Femur		Tibia		Fibula	
	<i>F</i>	<i>p</i>	<i>F</i>	<i>P</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Total sample												
Series	2.514	0.116	5.578	0.020	8.833	0.004	5.502	0.020	4.766	0.031	4.222	0.043
Sex	2.078	0.153	5.506	0.021	3.380	0.069	1.736	0.190	1.433	0.234	0.717	0.399
Subsample <2 years												
Series	0.019	0.892	0.212	0.648	1.742	0.196	0.367	0.547	0.000	0.993	0.082	0.777
Sex	0.036	0.851	0.708	0.405	0.537	0.469	0.125	0.725	0.001	0.979	0.055	0.821
Subsample ≥2 years												
Series	4.783	0.033	6.825	0.012	10.648	0.002	5.159	0.026	3.624	0.061	8.034	0.006
Sex	1.682	0.200	4.207	0.045	4.727	0.034	0.574	0.451	0.337	0.564	2.326	0.133

Results are shown for the entire combined sex sample (total) and for the subsamples that include individuals younger than 2 years of age (<2 years) and that includes individuals 2 years of age and older (≥2 years). The interaction between the effects of the series and of sex is not significant for any comparison

Table 4 Classical calibration models for each long bone length in the total sample, divided by sex and for the sexes combined

	Humerus	Radius	Ulna	Femur	Tibia	Fibula
Females						
<i>N</i>	41	40	40	58	54	38
Equation	Age = $\frac{\text{length}-79.94}{12.95}$	Age = $\frac{\text{length}-60.14}{9.22}$	Age = $\frac{\text{length}-67.04}{10.18}$	Age = $\frac{\text{length}-97.01}{19.58}$	Age = $\frac{\text{length}-80.43}{15.57}$	Age = $\frac{\text{length}-81.29}{15.24}$
MSE	1.37	1.36	1.34	1.21	1.25	1.39
<i>R</i> ²	0.90	0.91	0.92	0.93	0.93	0.92
<i>M</i>	131.6	98.5	115.2	180.4	149.4	165.4
<i>SD</i>	55.0	39.9	46.7	86.4	70.3	70.2
Min–max	55–225	44–168	50–211	61–351	54–302	51–292
Males						
<i>N</i>	61	65	54	82	75	55
Equation	Age = $\frac{\text{length}-79.66}{13.96}$	Age = $\frac{\text{length}-61.18}{10.12}$	Age = $\frac{\text{length}-67.61}{10.85}$	Age = $\frac{\text{length}-97.27}{20.99}$	Age = $\frac{\text{length}-80.58}{16.49}$	Age = $\frac{\text{length}-76.27}{16.51}$
MSE	0.95	1.08	1.10	0.92	1.00	1.03
<i>R</i> ²	0.94	0.93	0.92	0.94	0.93	0.94
<i>M</i>	136.9	103.6	113.6	183.2	151.5	157.5
<i>SD</i>	52.4	39.7	42.4	78.7	63.0	65.7
Min–max	52–256	44–202	50–206	59–369	52–308	49–298
Sexes combined						
<i>N</i>	102	105	94	140	129	93
Equation	Age = $\frac{\text{length}-79.97}{13.51}$	Age = $\frac{\text{length}-60.93}{9.74}$	Age = $\frac{\text{length}-67.64}{10.48}$	Age = $\frac{\text{length}-97.62}{20.28}$	Age = $\frac{\text{length}-80.84}{16.02}$	Age = $\frac{\text{length}-78.63}{15.90}$
MSE	1.13	1.21	1.21	1.06	1.11	1.17
<i>R</i> ²	0.92	0.91	0.92	0.93	0.93	0.93
<i>M</i>	134.8	101.7	114.3	182.0	150.6	160.7
<i>SD</i>	53.3	39.7	44.1	81.7	65.9	67.3
Min–max	52–256	44–202	50–114	59–369	52–308	49–298

Age is in years

MSE mean standard error (see text for more details) for the model, *R*² coefficient of determination from length regressed on age (length = $a \times \text{age} + b$), *M* mean long bone length (millimeters), *SD* standard deviation for long bone length (millimeters), *min–max* range of values for long bone length (millimeters)

Table 5 Classical calibration models for each long bone length in the subsample of individuals younger than 2 years of age (<2 years), divided by sex and for the sexes combined

	Humerus	Radius	Ulna	Femur	Tibia	Fibula
Females						
<i>N</i>	20	19	17	26	24	12
Equation	Age = $\frac{\text{length}-64.51}{27.54}$	Age = $\frac{\text{length}-50.25}{19.19}$	Age = $\frac{\text{length}-58.06}{18.66}$	Age = $\frac{\text{length}-75.20}{40.50}$	Age = $\frac{\text{length}-64.49}{30.41}$	Age = $\frac{\text{length}-58.93}{31.11}$
MSE	0.27	0.26	0.27	0.25	0.25	0.29
<i>R</i> ²	0.89	0.91	0.88	0.90	0.89	0.90
<i>M</i>	84.8	63.9	70.0	102.2	85.3	84.4
<i>SD</i>	21.6	15.3	13.7	30.1	21.3	25.8
Min–max	55–121	44–92	50–70	61–158	54–128	51–125
Males						
<i>N</i>	24	25	21	32	29	19
Equation	Age = $\frac{\text{length}-63.40}{28.93}$	Age = $\frac{\text{length}-50.06}{21.55}$	Age = $\frac{\text{length}-56.48}{22.91}$	Age = $\frac{\text{length}-72.70}{43.53}$	Age = $\frac{\text{length}-64.88}{29.88}$	Age = $\frac{\text{length}-55.13}{35.27}$
MSE	0.26	0.25	0.24	0.23	0.34	0.24
<i>R</i> ²	0.88	0.87	0.90	0.90	0.79	0.88
<i>M</i>	88.8	68.1	75.5	110.1	92.1	89.7
<i>SD</i>	20.4	14.5	16.5	30.6	21.5	23.8
Min–max	52–121	44–92	50–101	59–154	52–126	49–125
Sexes combined						
<i>N</i>	44	44	38	58	53	31
Equation	Age = $\frac{\text{length}-64.01}{28.22}$	Age = $\frac{\text{length}-50.29}{20.43}$	Age = $\frac{\text{length}-57.27}{21.17}$	Age = $\frac{\text{length}-74.04}{42.01}$	Age = $\frac{\text{length}-64.67}{30.13}$	Age = $\frac{\text{length}-57.13}{33.25}$
MSE	0.29	0.25	0.25	0.23	0.29	0.25
<i>R</i> ²	0.89	0.89	0.88	0.90	0.84	0.89
<i>M</i>	87.0	66.3	76.7	106.5	89.0	87.7
<i>SD</i>	20.8	14.8	15.3	30.4	21.6	24.3
Min–max	52–121	44–92	50–101	59–158	52–128	49–125

Age is in years

MSE mean standard error (see text for more details) for the model, *R*² coefficient of determination from length regressed on age (length=*a* × age + *b*), *M* mean long bone length (millimeters), *SD* standard deviation for long bone length (millimeters), *min–max* range of values for long bone length (millimeters)

includes only the younger children (<2 years). Figure 2 illustrates the raw residuals in the classical (A and B) and inverse calibration models (C and D). Whereas in the classical calibration model the raw residuals are scattered randomly about 0, in the inverse calibration models the residuals appear more spread out below the zero line.

The study sample also provided an accuracy test for the formulae published by Facchini and Veschi [23] and that by Rissech and coworkers [24–26]. Table 9 shows the accuracy results of Facchini and Veschi's [23] formulae, in which the length of the humerus bone provides the greatest accuracy, closely followed by the femur and the tibia. MR vary between −0.03 and 0.06, showing that there is only a slight overall overestimation. MR are consistently different from zero (*p* < 0.05) in the subsample which includes only the younger children (<2 years), but in the subsample of older children only the MR obtained from the length of the radius and ulna in males are significantly different from zero. MAR

is generally very low, varying between 0.04 and 0.12 years (under 2 months). Rissech and coworker's formulae (Table 10) are most accurate in the total sample when using the humerus, but in the subsample of younger children the tibia is most accurate and in the subsample of older children it is the femur. There is a general tendency for these formulae to underestimate age (MR varies between −1.51 and 0.58) which can be estimated within 0.74 to 1.58 years. MR show consistent differences from zero (*p* < 0.05) in the entire sample, but particularly for the length of the humerus and tibia.

Discussion

Previous approaches used in the estimation of age from the length of the diaphysis in long bones were either not devised specifically for this purpose, as is the case for tables of descriptive statistics from radiographic data, or show

Table 6 Classical calibration models for each long bone length in the subsample of individuals 2 years of age and older (≥ 2 years), divided by sex and for the sexes combined

	Humerus	Radius	Ulna	Femur	Tibia	Fibula
Females						
<i>N</i>	21	21	23	30	30	26
Equation	Age = $\frac{\text{length}-107.82}{9.65}$	Age = $\frac{\text{length}-77.71}{7.16}$	Age = $\frac{\text{length}-84.88}{8.22}$	Age = $\frac{\text{length}-131.28}{15.69}$	Age = $\frac{\text{length}-106.05}{12.74}$	Age = $\frac{\text{length}-106.63}{12.50}$
MSE	1.80	1.85	1.84	1.40	1.61	1.61
R^2	0.79	0.78	0.79	0.88	0.85	0.84
<i>M</i>	176.2	129.9	148.7	244.0	200.7	202.8
<i>SD</i>	36.4	26.9	31.4	60.8	50.5	48.9
Min–max	104–225	75–168	83–211	136–351	114–302	107–292
Males						
<i>N</i>	37	40	33	50	46	36
Equation	Age = $\frac{\text{length}-90.77}{12.49}$	Age = $\frac{\text{length}-66.59}{9.41}$	Age = $\frac{\text{length}-72.86}{10.14}$	Age = $\frac{\text{length}-115.2}{18.63}$	Age = $\frac{\text{length}-92.91}{14.91}$	Age = $\frac{\text{length}-86.82}{15.21}$
MSE	1.03	1.27	1.30	0.97	1.18	1.17
R^2	0.91	0.88	0.87	0.92	0.88	0.88
<i>M</i>	168.0	125.8	137.8	229.9	188.9	193.2
<i>SD</i>	42.1	33.9	35.4	62.6	50.3	50.7
Min–max	100–256	75–202	89–216	136–369	110–308	105–298
Sexes combined						
<i>N</i>	58	61	56	79	76	62
Equation	Age = $\frac{\text{length}-96.88}{11.38}$	Age = $\frac{\text{length}-70.31}{8.57}$	Age = $\frac{\text{length}-77.76}{9.27}$	Age = $\frac{\text{length}-122.31}{17.24}$	Age = $\frac{\text{length}-98.82}{13.88}$	Age = $\frac{\text{length}-95.96}{13.90}$
MSE	1.31	1.49	1.51	1.16	1.35	1.36
R^2	0.87	0.84	0.84	0.90	0.87	0.86
<i>M</i>	171.0	127.2	142.3	235.4	193.6	197.2
<i>SD</i>	40.0	31.5	33.9	61.9	50.4	49.8
Min–max	100–256	75–202	83–216	136–369	110–308	105–298

Age is in years

MSE mean standard error (see text for more details) for the model, R^2 coefficient of determination from length regressed on age (length= $a \times \text{age} + b$), *M* mean long bone length (millimeters), *SD* standard deviation for long bone length (millimeters), *min–max* range of values for long bone length (millimeters)

important statistical caveats that undermine the reporting and the accuracy of age estimates. In this study, a series of new regression methods are proposed for the estimation of age in remains of known or unknown sex from diaphyseal lengths of the long bones, using classical calibration, that address those specific concerns.

This study used samples from reference collections in Portugal and England, which were shown to differ in size. English children from both the Spitalfields and the St. Bride's collections lag behind the Lisbon children in growth. After the age of 2 years, children in the Lisbon collection have on average larger long bones than those from Spitalfields or St. Bride's, which is likely the result of differing social conditions during growth between the Portuguese and English samples. Despite these differences, the samples were combined in order to include more variation in the models and make them potentially applicable to a wider range of populations. Considering that it is

usually difficult or impossible to establish whether the model samples are representative of the growth status of unknown immature remains in a particular forensic case, or even in an archaeological sample, an approach that is not sample- or population-specific is likely to be more reliable. Under these circumstances, such an approach will fail less often in providing a reliable age estimate, but this estimate will have a larger confidence interval, due to sampling more variation.

Age estimation formulae were determined for the sexes separately and combined, in spite of the consistent similarities between males and females. In forensic casework, sex is required and age can be estimated with a slight increase in precision if sex is known. In an archaeological context, sex may not be determined prior to age estimation. Consequently, the development of sex-specific and sex-combined formulae was intended to address both forensic and archaeological applications.

Table 7 Accuracy results for classical calibration models in the total sample, and in the subsamples which include individuals younger than 2 years of age (<2 years) and individuals 2 years of age and older (≥2 years), divided by sex and for the sexes combined

	Humerus	Radius	Ulna	Femur	Tibia	Fibula
Total sample						
Females						
N	41	40	40	58	54	38
%Range	95.1	97.5	97.5	96.6	96.3	100
MAR	1.09	1.08	1.03	1.00	1.00	1.15
MR	0.00	0.00	0.00	0.00	0.00	0.00
Males						
N	61	65	54	82	75	55
%Range	98.4	96.9	100	98.8	98.7	100
MAR	0.80	0.88	0.91	0.77	0.82	0.85
MR	0.00	0.00	0.00	0.00	0.00	0.00
Sexes combines						
N	102	105	94	140	129	93
%Range	98.0	94.3	97.9	98.6	97.7	98.9
MAR	0.93	0.98	0.99	0.88	0.91	0.99
MR	0.00	0.00	0.00	0.00	0.00	0.00
Subsample of <2 years						
Females						
N	20	19	17	26	24	12
%Range	95.0	94.7	94.1	96.2	95.8	91.7
MAR	0.16	0.16	0.18	0.16	0.16	0.17
MR	0.00	0.00	0.00	0.00	0.00	0.00
Males						
N	24	25	21	32	29	19
%Range	100	96.0	95.2	93.8	96.6	94.7
MAR	0.19	0.18	0.18	0.17	0.25	0.16
MR	0.00	0.00	0.00	0.00	0.00	0.00
Sexes combines						
N	44	44	38	58	53	31
%Range	95.5	95.5	97.4	94.8	96.2	93.5
MAR	0.17	0.18	0.19	0.16	0.21	0.17
MR	0.00	0.00	0.00	0.00	0.00	0.00
Subsample of ≥2 years						
Females						
N	21	21	23	32	30	26
%Range	100	100	100	93.8	93.3	100
MAR	1.33	1.37	1.38	1.05	1.22	1.29
MR	0.00	0.00	0.00	0.00	0.00	0.00
Males						
N	37	40	33	50	46	36
%Range	94.6	100	100	92.0	97.8	100
MAR	0.83	0.98	1.02	0.74	0.93	0.94
MR	0.00	0.00	0.00	0.00	0.00	0.00
Sexes combined						
N	58	61	56	82	76	62
%Range	98.3	98.4	98.2	95.1	97.4	98.4

Table 7 (continued)

	Humerus	Radius	Ulna	Femur	Tibia	Fibula
MAR	1.02	1.13	1.18	0.89	1.05	1.07
MR	0.00	0.00	0.00	0.00	0.00	0.00

Differences (MAR and MR) are in years

The division of the sample at the age of 2 enables use of the most appropriate formulae for each age group, as these accommodate biological differences in the growth process. These differences include faster bone growth (regression slope) and reduced individual variation (regression error) in children under the age of 2 years. Although the calibration models for the total sample can be used for the entire prepubertal age range (0–12 years), the models for the two subsamples (<2 and ≥2 years) provide estimates that are more representative of the age-related differences in growth and more accurate for younger individuals.

The use of classical calibration over inverse calibration finds considerable support from the accuracy tests performed here. The classical calibration models have no bias (MR are zero), when compared with the inverse calibration models (MR different from zero). In addition, the classical calibration models shown here are not only similarly accurate but show no loss of efficiency, despite the expected larger variability for classical calibration when compared with inverse calibration [32, 33]. In fact, the mean of absolute residuals suggest that both models are equally efficient.

One or more of the formulae in Tables 4, 5, and 6 can be used to estimate the age of unknown immature skeletal remains from diaphyseal long bone length, according to the information available for bone length, age group, or sex. If, for example, an unknown individual has a femur measuring 190 mm in length and a nonsex-specific estimate is required, the femur length formula in Table 4 can be used to estimate age. Femur length is substituted into the formula $age = (length - 97.62) / 20.28$, thus obtaining an age estimate of 4.56 years after performing the calculation $(190 - 97.62) / 20.28$. The 95 % confidence interval for age can be calculated using the MSE (1.06) for the formula. The confidence interval is obtained first by multiplying the MSE by 2, the critical value in the *t* distribution that includes 95 % of the population ($1.06 \times 2 = 2.12$). This value is then added to and subtracted from the age estimate (4.56 ± 2.12 years) to obtain a 95 % confidence interval for the estimated age, which in this case is between 2.44 and 6.68 years. However, these confidence intervals in the classical calibration models cannot be described as there is a 95 % probability that the age falls between 2.44 and 6.68 years, but instead that there is a 95 % confidence that the age lies between 2.44 and 6.68 [35]. It is also important to highlight that the formula provided here are for use in dry bone, as there may be a certain amount of shrinkage

Table 8 Accuracy results for inverse calibration models in the total sample, and in the subsamples which include individuals younger than 2 years of age (< 2 years) and individuals 2 years of age and older (≥ 2 years), divided by sex and for the sexes combined

	Humerus	Radius	Ulna	Femur	Tibia	Fibula
Total sample						
Females						
N	41	40	40	58	54	38
%Range	95.1	95.0	95.0	93.1	96.3	97.4
MAR	1.02	0.98	0.97	1.01	0.94	1.10
MR	0.06	0.16	-0.05	0.47*	0.07	0.00
Males						
N	61	65	54	82	75	55
%Range	100	93.8	96.3	78.0	94.7	92.7
MAR	0.85	0.84	0.99	1.16	0.93	0.95
MR	0.38*	-0.16	0.56*	0.95*	0.52*	0.52*
Sexes combined						
N	102	105	94	140	129	93
%Range	98.0	93.3	95.7	88.7	98.4	95.7
MAR	0.91	0.92	0.96	1.08	0.92	0.99
MR	0.26*	-0.38*	0.27*	0.73*	0.31*	0.29*
Sample < 2 years						
Females						
N	20	19	17	26	24	13
%Range	90.0	89.5	94.1	100	95.8	91.7
MAR	0.22	0.25	0.28	0.23	0.17	0.22
MR	-0.20*	0.17*	0.21*	-0.22*	0.06	0.10
Males						
N	24	25	21	32	29	19
%Range	100	96.0	95.2	90.6	82.8	57.9
MAR	0.17	0.18	0.18	0.17	0.40	0.45
MR	-0.03	-0.03	0.06	-0.08*	0.33*	0.44*
Sexes combined						
N	21	21	23	32	30	26
%Range	97.7	88.6	94.7	93.1	92.5	90.3
MAR	0.18	0.24	0.19	0.19	0.28	0.35
MR	-0.11*	-0.22*	-0.12*	-0.15*	0.19*	0.29*
Sample ≥ 2 years						
Females						
N	21	21	23	32	30	26
%Range	95.2	95.2	100	84.4	93.3	96.2
MAR	1.21	1.18	1.42	1.31	1.28	1.24
MR	-0.38	0.12	0.61	0.99*	0.74*	0.51
Males						
N	37	40	33	50	46	36
%Range	91.9	92.5	97.0	96.0	100	100
MAR	0.84	0.97	1.14	0.76	0.92	0.97
MR	-0.50*	-0.48*	0.58*	0.15	0.15	0.36*
Sexes combined						
N	58	61	56	82	76	62
%Range	93.1	96.7	96.4	82.9	94.7	95.2

Table 8 (continued)

	Humerus	Radius	Ulna	Femur	Tibia	Fibula
MAR	1.16	1.11	1.12	0.90	1.05	1.04
MR	0.68*	0.24	-0.04	-0.49*	-0.44*	-0.39*

Differences (MAR and MR) are in years

* $p < 0.05$

relative to wet bone, which is more significant in the early ages [42].

Although testing the accuracy of the regression models in the same sample that was used to develop them may be considered inappropriate, the purpose here was not to carry out an independent test of the models but rather a test of their comparative performance. In fact, one would expect the models to perform well in the same sample that was used to develop them, but the reality is that the expectation was only met for the classical calibration model. This confirms the notion that inverse calibration models are inherently biased and should not be used to develop age estimation methods. This has already been widely acknowledged in the adult age estimation literature [32–34], where the systematic bias seen in most adult age estimation methods is considered to be a direct consequence of the use of least squares regression and inverse calibration (age regressed on skeletal age indicator). Consequently, it is perhaps time to acknowledge this same caveat when estimating age of non-adults. Although Aykroyd and coworkers [32] assert that systematic bias in inverse calibration models is reduced as the correlation between age and the skeletal indicator increases, the relatively high correlation between age and long bone length during growth was not sufficient to prevent significant biases. Using the sex combined models for the total sample (Table 8), the inverse calibration formulae yields age estimates with an average bias of about 3 months (0.25 years), whereas the classical calibration formulae yields age estimates with an average bias of 0 months that is with no bias. With the same formulae, an average of 94.9 % of the individuals will have their chronological age included in the confidence interval, compared with an average of 97.4 % using the classical calibration formulae. Although differences between the classical and inverse calibration models may seem small, the difference is significant in that use of classical calibration reduces the risk of misidentification of unknown individuals based on their age. The accuracy tests for the classical calibration model also show that the mean standard error can be reliably used to calculate 95 % confidence intervals. For most age estimation formulae, 95 % or more of the children had their true chronological age included in the confidence interval.

When the classical calibration models are examined in more detail, it is clear that age can be estimated with more accuracy and efficiency in the subsample that includes

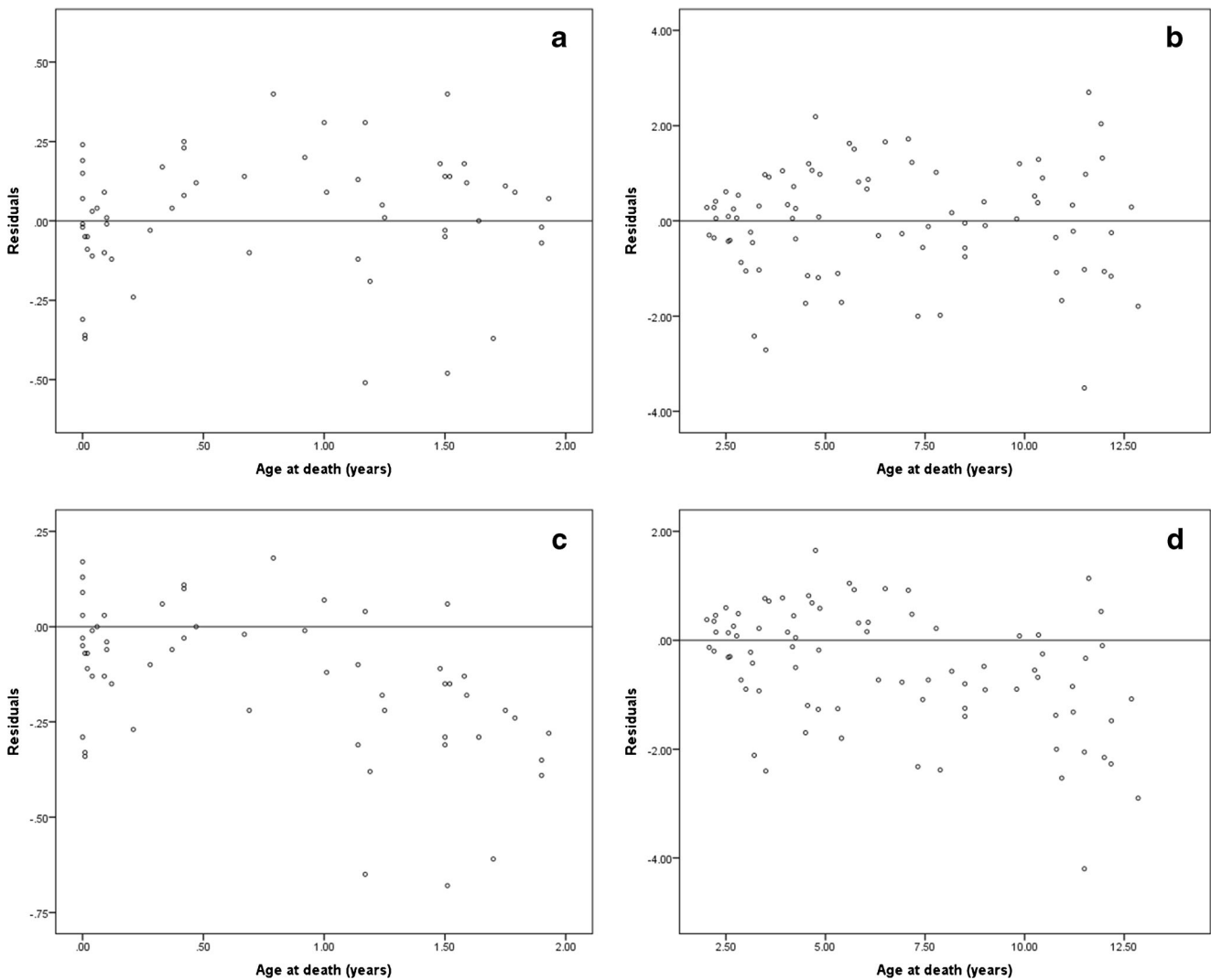


Fig. 2 Scatterplot of the raw residuals in the classical calibration (**a** children under 2 years of age and **b** children 2 years of age and older) and inverse calibration (**c** children under 2 years of age and **d** children

2 years of age and older) models for the relationship between long bone length (femur) and age, in the combined sex sample

individuals under the age of 2 years. This is expected as variation in growth is smaller in this age group and hence MAR will also be smaller when compared with the subsample that includes only individuals aged 2 years and older. Conversely, the average percentage of individuals whose chronological age is included in the confidence interval is actually slightly smaller in the <2-year subsample (94.6 %), compared with the ≥ 2 -year subsample (96.9 %). In general, the classical calibration model is most efficient for males, although it seems slightly more efficient for females in the subsample of children under 2 years of age. This may be related to the fact that the male and females sample size is more even under the age of 2, but the older female sample is smaller than that of males such that the female model is more likely to be influenced by random fluctuation in size. In fact, females show more variation as demonstrated by overall larger standard deviation

about the mean. The diaphyseal length of the femur is the variable with the least amount of expected error, averaging around 1.06 years (MSE) about the mean estimated age in the total sample with the sexes combined, and the least amount of real error, averaging around 0.88 years (MAR). By contrast, the diaphyseal length of the ulna shows the largest amount of expected (1.21 years) and real error (0.99 years). For the sample of < 2 years, the length of the tibia is the least suitable for age estimation.

The greatest limitation of this study is the size of the samples and their age distribution. Although it would be preferable to have a more evenly balanced age distribution, and greater representation of older children, it is currently impossible to add more identified skeletal material. This study combines two of the largest series of documented immature skeletal remains, and includes slightly more individuals,

Table 9 Accuracy of Facchini and Veschi's (2004) regression formulae in the total sample, and in the subsamples which include individuals younger than 2 years of age (< 2 years) and individuals 2 years of age and older (≥ 2 years), divided by sex and for the sexes combined

	Humerus	Radius	Ulna	Femur	Tibia	Fibula
Total sample						
Females						
N	41	40	40	58	54	38
MAR	0.08	0.08	0.08	0.08	0.08	0.09
MR	0.01	0.02	0.01	0.00	0.01	0.01
Males						
N	61	65	54	82	75	55
MAR	0.07	0.08	0.08	0.07	0.07	0.07
MR	0.02	0.05*	0.04*	0.02*	0.03*	0.03*
Sexes combines						
N	102	105	94	140	129	93
MAR	0.07	0.08	0.08	0.07	0.07	0.08
MR	0.01	0.04*	0.03*	0.01	0.02*	0.02*
Sample of <2 years						
Females						
N	20	19	17	26	24	12
MAR	0.05	0.05	0.04	0.05	0.05	0.05
MR	0.03*	0.03*	0.03*	0.03	0.03*	0.04
Males						
N	24	25	21	32	29	19
MAR	0.05	0.07	0.07	0.06	0.06	0.06
MR	0.04*	0.06*	0.05*	0.04*	0.04*	0.05*
Sexes combines						
N	44	44	38	58	53	31
MAR	0.05	0.06	0.06	0.05	0.05	0.06
MR	0.03*	0.05*	0.04*	0.04*	0.04*	0.04*
Sample of ≥ 2 years						
Females						
N	21	21	23	32	30	26
MAR	0.12	0.11	0.11	0.11	0.11	0.11
MR	-0.03	0.00	0.00	-0.02	-0.01	0.00
Males						
N	37	40	33	50	46	36
MAR	0.07	0.09	0.09	0.08	0.08	0.08
MR	0.01	0.05*	0.03*	0.01	0.02	0.01
Sexes combines						
N	58	61	56	82	76	62
MAR	0.09	0.10	0.10	0.09	0.09	0.09
MR	-0.01	0.03*	0.02	0.00	0.00	0.01

Differences (MAR and MR) are in years

* $p < 0.05$

particularly males, than Facchini and Veschi [23] and about twice as many as that of Rissech and coworkers [24–26]. A larger sample, particularly one which included a larger

proportion of older children, would include more variation and possibly reduce the amount of expected error. Another limitation of this study was the lack of opportunity to provide an independent test of the accuracy of the classical calibration models in view of the scarcity of identified immature human skeletal material. It is hoped that with access to similar collections by the authors and/or by other researchers will provide a much desired independent test.

The efficiency of the formulae provided in this study is difficult to assess relative to that of other published formulae [23–26], because previous studies have only used inverse calibration for modeling age and long bone length, and only Rissech and coworkers [25, 26] have actually provided error rates for their formulae. The comparison is further complicated because error rates provided by Rissech and coworkers [25, 26] cannot be directly compared with those of this study, because only an approximation of the standard error can be obtained from the classical calibration models. The standard error of the estimate provided by Rissech and coworkers [25, 26] for estimating age from the diaphyseal length of the humerus and of the tibia is 1.399 and 1.777 years respectively. The mean standard error in the formulae for the same models in the study sample, but using classical calibration, are 1.13 and 1.11 years for the humerus and tibia, respectively. In the end, it is difficult to determine which error estimates are actually greater, because of the issues discussed above, but also due to Rissech and coworker's sample including individuals up to 19 and 15 years of age, which increases the amount of error by introducing increased age variation.

Considering that Facchini and Veschi's [23] and Rissech and coworkers' [24–26] recently published formulae for age estimation from long bone lengths, were based on least squares and inverse calibration, the overall expectation is for these models to provide biased age estimates. Overall, Facchini and Veschi's [23] formulae seem to outperform those of Rissech and coworkers [24–26]. This may suggest that the sample used by Facchini and Veschi [23] includes children of approximately the same size for age or only slightly larger than children in this study sample. Conversely, children in Rissech and coworkers' [24–26] sample seem smaller for their age. However, this may be an artifact of the bias introduced by inverse calibration. In fact, one would expect greater similarities in size for age in Rissech and coworkers' [24–26] sample, given that the some individuals from the same collections were used. Conversely, the difference may also be explained by the fact that Rissech and coworkers' [24–26] sample includes children from the Coimbra collection and these may be overall smaller in size for age, compared with the Lisbon, Spitalfields, and St. Bride's collections. The bias introduced by inverse calibration may also explain why Facchini and Veschi's [23] formula are more accurate, when in fact the average long bone length in their sample seems slightly greater than that in the study sample. The results of the tests of

Table 10 Accuracy of Rissech and coworker’s (2008, 2012) [26] formulae in the total sample, and in the subsamples which include individuals younger than 2 years of age (<2 years) and individuals 2 years of age and older (≥2 years), divided by sex and for the sexes combined

	Humerus	Femur	Tibia
Total sample			
Females			
N	41	58	54
%Range	95.1	–	100
MAR	1.17	1.26	1.10
MR	–0.20	–0.57*	0.04
Males			
N	61	82	75
%Range	100	–	100
MAR	0.98	0.96	0.98
MR	0.11	–0.54*	–0.32*
Sexes combines			
N	102	140	129
%Range	98.0	–	100
MAR	1.05	1.09	1.05
MR	–0.01	–0.46*	–0.67*
Sample of <2 years			
Females			
N	20	26	24
%Range	100	–	100
MAR	1.11	1.58	0.93
MR	–0.74*	–1.51*	–0.64*
Males			
N	24	32	29
%Range	100	–	100
MAR	0.97	1.30	1.06
MR	–0.56*	–1.25*	–0.91*
Sexes combines			
N	44	58	53
%Range	100	–	100
MAR	1.03	1.38	0.86
MR	–0.64*	–1.30*	–0.59*
Sample of ≥2 years			
Females			
N	21	32	30
%Range	90.5	–	100
MAR	1.22	0.99	1.24
MR	0.31	0.20	0.58*
Males			
N	37	50	46
%Range	100	–	100
MAR	0.98	0.74	0.93
MR	0.55*	–0.08	0.05
Sexes combines			
N	58	82	76
%Range	96.6	–	100

Table 10 (continued)

	Humerus	Femur	Tibia
MAR	1.07	0.88	1.07
MR	0.46*	0.13	0.29*

Differences (MAR and MR) are in years

**p* < 0.05

Facchini and Veschi’s [23] and Rissech and coworkers’ [24–26] formulae illustrates the effects of using inverse calibration for age estimation, where a method developed on a more closely related sample will be less reliable than a method developed on another sample that reflects differences in size. These biases caution against using population-specific methods, particularly when they were developed using inverse calibration. In addition to the formulae provided by Facchini and Veschi [23], Rissech and coworkers [24–26] published techniques for fetal age estimation from measurements of the long bones based on inverse calibration [43–45] are also likely to be equally biased.

A key consideration is the applicability of the age estimation formulae presented here to the analysis of human immature skeletal remains from forensic cases. Numerous studies have shown that long bone growth rates of children in past populations or in populations from developing countries which experience poor nutrition or increased risk of disease are lower than those in modern Western industrialized nations [46]. Skeletal growth profiles in the Spitalfields, St. Bride’s, and Lisbon samples show consistently slower rates of long bone growth compared with the modern reference established by the Denver Growth Study [4, 46]. In addition, secular changes have had a very strong impact on body size, particularly in the last 100 years in Europe, where height in adults and children increased steadily due to improvements in living conditions [47].

The Lisbon collection children lived during a period when most European children were experiencing increased growth in height compared with previous generations [47, 48]. However, Portuguese children did not experience the major secular increase in height until after 1960. While the height of Portuguese children increased by only 0.4–2.6 cm between 1906 and 1936, there was an increase of 10.1–15.9 cm between 1966 and 2006 [49]. The Lisbon collection children predate these major changes in height and were on average shorter than present-day Portuguese children of the same age. English children also experienced a similar increase in height over the twentieth century [50], but the Spitalfields and St. Bride’s children are from the eighteenth and nineteenth centuries. Consequently, all children in the study sample predate the secular increase in height documented for European children during the twentieth century and are more representative of populations experiencing lower levels of social and economic

development, including archaeological populations and children from developing countries.

These findings have obvious implications for the use of long bone length as an indicator of age. As the study sample represents a group of children who experienced moderate to severe malnutrition and were exposed to high disease loads, they are stunted in growth [4, 46]. Consequently, the formulae provided here are unlikely to be useful in modern medicolegal contexts of the developed world, as they will not reflect the current growth status of children in most developed nations. The study sample may be considered representative of children in present-day populations which have poor nutrition and increased risk of disease, including those from developing countries and some more deprived communities in developed nations. Therefore, the age estimation formulae should only be used for estimating age in children who experienced environmental conditions similar to those of Portugal and England living in urban environments between 250 and 50 years ago. Although it may be impossible to determine whether the formulae are suitable to a specific population or group, the range and mean long bone lengths provided here may offer a crude gauge to make such an assessment. The fact that the growth of the long bones is more similar between Portuguese and English children in the study sample before the age of 2 years, suggests that the formulae for children under 2 years of age may be more suitable for forensic purposes and for a wider range of populations. Differences in growth due to environmental influences are usually established by the age of 2 years [47], but become increasingly noticeable after that age because of the cumulative effects of environmental insults on growth. Whereas regression formulae provided in this study are more likely to be useful in forensic investigations involving immature remains from developed nations, particularly if the child is under 2 years of age, there is little to turn to for help when estimating age of child remains in the developed world. If the expert is willing to sacrifice the error rate of the age estimate, then some of the tables [9–14] and graphs [21, 22] mentioned before can provide some guidance.

A further concern relating to the use of formulae derived from skeletal size in deceased children for age estimation is mortality bias. The growth of children who die prematurely may not be representative of their living counterparts as children who experience stunted growth also have a greater probability of dying. As a result, skeletal samples tend to include a disproportionate number of stunted children [51]. Consequently, the formulae provided here may reflect an overestimation of the growth deficit in an already deprived population. This will be less of a concern in archaeological remains, since in this case children will also have died prematurely from natural causes, whereas child remains in forensic investigations are found more commonly in circumstances where a violent death has occurred.

For reasons similar to those discussed above, Facchini and Veschi's [23] and Rissech and coworkers' [24–26] formulae will not provide reliable estimates of age in modern immature skeletal remains from Western Europe, and in particular from the Iberian Peninsula as argued by Rissech and coworkers [25]. Rissech and coworkers [25] assert that their formulae are valid for forensic age estimation in modern children because they did not detect delayed growth in their series compared with modern growth models. They further argued that the homogeneity observed in the maximum length of the adult femur, between their series and the Spanish documented skeletal collection of the Universidad Autonoma de Barcelona (UAB) [24] implied a lack of growth deficit in their subadult series. However the adults in the UAB collection were born between 1892 and 1959 [24], making them of the same birth cohort of the Lisbon collection children. As a result, they are likely to have experienced similar deficit during growth as the Lisbon sample, resulting in smaller adult size than would have been attained in modern adults. Given the secular effects discussed above, European children are considerably taller today than they were 50 or 100 years ago. Spanish children, in particular, are among the Europeans who have experienced the greatest secular increase in height since the middle of the twentieth century, about 2.4 cm/decade [48, 52]. Unfortunately, Rissech and coworkers' [24–26] have incorrectly asserted their samples as appropriate for use in a modern Iberian or Western European forensic context. This provides a cautionary note against assumptions of temporal continuity in skeletal growth in populations when using cemetery collections of identified human skeletons to represent present-day populations, as secular trends must be considered.

Conclusions

The long bone lengths collected in this study are derived from one of the largest assemblages of known sex and age immature skeletons presently available, not only in terms of overall size but because the entire age range is reasonably well represented. Analysis of these data demonstrates that inverse calibration is an inappropriate approach for developing age estimation formulae and that it should be abandoned when modeling age and a skeletal indicator of age, even in immature skeletal remains. The classical calibration models presented here provide a series of new formulae for age estimation from the diaphyseal length of the long bones that can be potentially used in a variety of contexts. The long bone growth rates in the study sample are lower than those of the modern Western industrialized nations which have undergone considerable secular change because of improved living standards. As a result, the formulae cannot be reliably applied to forensic cases involving recent children, but can possibly provide the best available estimates when children from developing

countries are involved or children from poor communities in developed nations. The formulae can also be used in archaeological populations where long bone growth rates in children are similar those in England in the eighteenth to nineteenth centuries and Portugal in the early twentieth century. Previously published formulae for age estimation from long bone lengths have provided unsatisfying results due to failure to consider secular trends and inappropriate statistical treatment of data.

Acknowledgments We would like to thank the three anonymous reviewers for their valuable comments and suggestions that improved the quality of the initial manuscript. Data collection and analysis for this study were partially supported by FCT grants # SFRH/BD/4917/2001 and SFRH/BPD/22142/2005 and the SYNTHESYS program.

References

1. Ubelaker DH (1987) Estimating age at death from immature human skeletons: an overview. *J Forensic Sci* 32(5):1254–1263
2. Lewis AB, Garn SM (1960) The relationship between tooth formation and other maturational factors. *Angle Orthod* 30:70–77
3. Lewis M, Flavel A (2006) Age assessment of child skeletal remains in forensic contexts. In: Schmitt A, Cunha E, Pinheiro J (eds) *Forensic anthropology and medicine: complementary sciences from recovery to cause of death*. Humana Press, New Jersey, pp 243–257
4. Cardoso HFV (2007) Environmental effects on skeletal versus dental development: using a documented subadult skeletal sample to test a basic assumption in human osteological research. *Am J Phys Anthropol* 132:223–233
5. Ferembach D, Schwidetzky I, Stloukal M (1980) Recommendations for age and sex diagnoses of skeletons. *J Hum Evol* 9:517–549
6. Rösing FW, Graw M, Marré B, Ritz-Timme S, Rothschild MA, Röttscher K, Schmeling A, Schröder I, Geserick G (2007) Recommendations for the forensic diagnosis of sex and age from skeletons. *Homo* 58:75–89
7. Cunha E, Baccino E, Martrille L, Ramsthaler F, Prieto J, Schuliar Y, Lynnerup N, Cattaneo C (2009) The problem of aging human remains and living individuals: a review. *Forensic Sci Int* 193:1–13
8. Franklin D (2010) Forensic age estimation in human skeletal remains: current concepts and future directions. *Legal Med* 12:1–7
9. Maresh MM (1943) Growth of major long bones in healthy children. A preliminary report on successive roentgenograms of the extremities from early infancy to twelve years of age. *Am J Dis Child* 66:227–257
10. Maresh MM (1955) Linear growth of the long bones of extremities from infancy through adolescence. *Am J Dis Child* 89:725–742
11. Maresh MM (1970) Measurements from roentgenograms. In: McCannon RW (ed) *Human growth and development*. Springfield, Illinois, pp 157–188
12. Ghantus MK (1951) Growth of the shaft of the human radius and ulna during the first two years of life. *Am Hum Biol* 12:129–148
13. Anderson M, Messener MB, Green WT (1964) Distribution of lengths of the normal femur and tibia in children from one to eighteen years of age. *J Bone Joint Surg* 46-A:1197–202
14. Gindhart PS (1973) Growth standards for the tibia and radius in children aged one month through eighteen years. *Am J Phys Anthropol* 39:41–48
15. Johnston FE (1962) Growth of the long bones of infants and young children at Indian Knoll. *Am J Phys Anthropol* 20:249–254
16. Walker PL (1969) The linear growth of the long bones in late woodland Indian children. *Proc Indiana Acad Sci* 78:83–87
17. Stloukal M, Hanáková H (1978) Die Länge der Längsknochen altslawischer Bevölkerungen unter besonderer Berücksichtigung von Wachstumsfragen. *Homo* 29:53–69
18. Sundick RI (1978) Human skeletal growth and age determination. *Homo* 29:228–249
19. Hoppa RD (1992) Evaluating human skeletal growth: an Anglo-Saxon example. *Int J Osteoarchaeol* 2:275–288
20. Fazekas IG, Kosa F (1978) *Forensic fetal osteology*. Akademiai Kiado, Budapest
21. Stewart TD (1976) Identification by the skeletal structures. In: Camps FE (ed) *Gradwohl's legal medicine*. Wright, Bristol, pp 123–154
22. Hoffman JM (1979) Age estimations from diaphyseal lengths: two months to twelve years. *J Forensic Sci* 24:461–469
23. Facchini F, Veschi S (2004) Age determination on long bones in a skeletal subadults samples (*b*–12 years). *Coll Anthropol* 1:89–98
24. Rissech C, Schaefer M, Malgosa A (2008) Development of the femur—implications for age and sex determination. *Forensic Sci Int* 180:1–9
25. Rissech C, López-Costa O, Turbón D (2013) Humeral development from neonatal period to skeletal maturity—application in age and sex assessment. *Int J Legal Med* 127:201–212
26. López-Costas O, Rissech C, Tranco G, Turbón D (2012) Postnatal ontogenesis of the tibia. Implications for age and sex estimation. *Forensic Sci Int* 214:207.e1–11
27. Danforth ME, Wrobel GD, Armstrong CW, Swanson D (2009) Juvenile age estimation using diaphyseal long bone lengths among ancient Maya populations. *Soc Am Arch* 20:3–13
28. Boccone S, Cremasco MM, Bortoluzzi S, Moggi-Cecchi J, Massa ER (2010) Age estimation in subadult Egyptian remains. *Homo* 61:337–358
29. Primeau C, Friis L, Sejrsen B, Lynnerup N (2012) A method for estimating age of Danish medieval sub-adults based on long bone length. *Antrop Anz* 69:317–333
30. Smith SL, Buschang PH (2005) Longitudinal models of long bone growth during adolescence. *Am J Hum Biol* 17:731–745
31. Smith SL, Buschang PH (2004) Variation in longitudinal diaphyseal long bone growth in children three to ten years of age. *Am J Hum Biol* 16:648–657
32. Aykroyd RG, Lucy D, Pollard AM, Solheim T (1997) Technical note: regression analysis in adult age estimation. *Am J Phys Anthropol* 104:259–265
33. Konigsberg LW, Frankenberg SR, Walker RB (1994) Regress what on what?: paleodemographic age estimation as a calibration problem. In: Paine RR (ed) *Integrating archaeological demography: multidisciplinary approaches to prehistoric population*. Southern Illinois University, Carbondale, pp 64–88
34. Lucy D, Pollard A (1995) Further comments on the estimation of error associated with the Gustafson dental age estimation method. *J Forensic Sci* 40:222–227
35. Lucy D (2005) *Introduction to statistics for forensic scientists*. Wiley, England, pp 75–93
36. Cardoso HFV (2006) Brief communication: the collection of identified human skeletons housed at the Bocage museum (National Museum of Natural History), Lisbon, Portugal. *Am J Phys Anthropol* 129:173–176
37. Molleson T, Cox M, Waldron AH, Whittaker DK (1993) *The Spitalfields project, Vol. 2: the anthropology: the middling sort*. Council for British Archaeology, York
38. Scheuer JL, Bowman JE (1995) Correlation of documentary and skeletal evidence in the St. Bride's crypt population. In: Saunders SR, Herring A (eds) *Grave reflections: portraying the past through cemetery studies*. Canadian Scholars' Press, Toronto, pp 49–70
39. Buikstra J, Ubelaker D (1994) Standards for data collection from human skeletal remains. *Arkansas Archeological Survey, Arkansas*

40. Ulijaszek SJ, Kerr DA (1999) Anthropometric measurement error and the assessment of nutritional status. *Br J Nutr* 82:165–177
41. Humphrey LT (2003) Linear growth variation in the archaeological record. In: Thompson JL, Krovitz GE and Nelson AJ (eds) *Patterns of growth variation in the genus Homo*. Cambridge University Press, Cambridge, pp 144–169
42. Huxley AK (1998) Analysis of shrinkage in human fetal diaphyseal lengths from fresh to dry bone using Petersohn and Kohler's data. *J Forensic Sci* 43:423–426
43. Scheuer J, Musgrave J, Evans S (1980) The estimation of late fetal and perinatal age from limb bone length by linear and logarithmic regression. *Ann Hum Biol* 7:257–65
44. Sherwood RJ, Meindl RS, Robinson HB, May RL (2000) Fetal age: methods of estimation and effects of pathology. *Am J Phys Anthropol* 113:305–315
45. Carneiro C, Curate F, Borralho P, Cunha E (2013) Radiographic fetal osteometry: approach on age estimation for the portuguese population. *Forensic Sci Int*. doi:10.1016/j.forsciint.2013.05.039
46. Humphrey L (2000) Growth studies of past populations: an overview and an example. In: Cox M, Mays S (eds) *Human osteology in archaeology and forensic sciences*. Greenwich Medical Media Ltd, London, pp 23–38
47. Cole TJ (2000) Secular trends in growth. *Proc Nutr Soc* 59:317–324
48. Susanne C, Bodzsar E, Bielicki T, Hauspie B, Hulanicka B, Lepage Y, Rebato E, Vercauteren M (2001) Changements séculaires de la croissance et du développement en Europe. *Antropo* 0:71–90
49. Cardoso HFV (2008) Secular changes in body height and weight of portuguese boys over one century. *Am J Hum Biol* 20:270–277
50. Cameron N (1979) The growth of London school children 1904–1966: an analysis of secular trend and intra-county variation. *Ann Hum Biol* 6:505–525
51. Saunders SR, Hoppa RD (1993) Growth deficit in survivors and non-survivors: biological mortality bias in subadult skeletal samples. *Phys Anthropol* 36:127–151
52. Rebato E (1998) The studies on secular trend in Spain: a review. In: Bodzsar E, Susanne C (eds) *Secular growth changes in Europe*. Eötvös University Press, Budapest, pp 279–317