Archaeol Anthropol Sci (2016) 8:731–750
DOI 10.1007/s12520-015-0252-5

ORIGINAL PAPER

Estimating body mass from skeletal material: new predictive equations and methodological insights from analyses of a known-mass sample of humans

Marina Elliott[1,2] · **Helen Kurki**[3] · **Darlene A. Weston**[4,5] · **Mark Collard**[1,6]

Received: 22 December 2014 / Accepted: 12 May 2015 / Published online: 16 June 2015

Abstract Estimating body mass from skeletal material is a key task for many biological anthropologists. As a result, several sets of regression equations have been derived for cranial and postcranial material. The equations have been applied to a wide range of specimens, but several factors suggest they may not be as reliable as generally assumed. Specifically, since many of the equations were derived from small reference samples using proxies for key variables and/or mean data, the nature of the relationship between the skeletal variables and body mass has often not been adequately demonstrated. In addition, few of the equations have been validated on known samples, making their accuracy and precision uncertain. Lastly, because no study has used cranial and postcranial material from the same individuals, the two approaches have never been systematically compared. The present study responded to these issues by deriving new regression equations from cranial and postcranial material using a large sample of modern humans of known-mass and associated skeletal variables measured from CT data. The equations were then tested on an independent sample, also of known mass. The results show that the newly derived equations estimate mass more accurately than existing equations for most variables. However, improvements were modest and accuracy rates remained relatively low. In addition, variables that had previously been argued to be ideal predictors were not the most accurate, and the current criteria used to assess equations did not ensure reliability. Overall, the results suggest that body mass estimates must be used cautiously and that further research is required.

Keywords Biological anthropology · Bioarchaeology · Fossil hominin · Osteology · Palaeoanthropology · Forensic anthropology

Electronic supplementary material The online version of this article (doi:10.1007/s12520-015-0252-5) contains supplementary material, which is available to authorized users.

✉ Mark Collard
mcollard@sfu.ca

[1] Human Evolutionary Studies Program and Department of Archaeology, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada

[2] Evolutionary Studies Institute, University of the Witwatersrand, Johannesburg, South Africa

[3] Department of Anthropology, University of Victoria, Victoria, BC V8W 2Y2, Canada

[4] Department of Anthropology, University of British Columbia, Vancouver, BC V6T 1Z1, Canada

[5] Department of Human Evolution, Max Planck Institute of Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany

[6] Department of Archaeology, University of Aberdeen, St Mary's Building, Aberdeen AB24 3UF, UK

Introduction

Bone-derived estimates of body mass play an important role in biological anthropology. Such estimates are crucial for understanding the evolutionary history of humans and other hominins (Damuth and MacFadden 1990; Delson et al. 2000). They are required to understand the adaptive strategies of hominin species and to accurately compare features across fossil groups (Frayer 1984; Wood and Collard 1999; Delson et al. 2000; DeSilva 2011). Body mass estimates are also necessary for interpreting the characteristics of individuals from more recent periods, particularly in terms of growth and

development, and health (Steckel and Rose 2002; Cohen and Crane-Kramer 2007). Lastly, because body mass is a conspicuous individualizing feature and a potential influence on taphonomic processes, body mass estimation is increasingly being used in forensic research (Suskewicz 2004; Rainwater et al. 2007; Agostini and Ross 2011; Byard 2012).

Over the last quarter of a century, regression analysis has been the main method used to estimate body mass from skeletal remains. The most common approach has been called "inverse calibration" (Konigsberg et al 1998) and involves the regression of body mass on a skeletal variable in samples of extant taxa to generate an equation with which to predict mass in an unknown specimen (Ruff et al. 2012). The resulting equations take the form $Y=a+bX$, where Y is the estimated mass, X is the skeletal measurement, a is the intercept of the regression line, and b is its slope. The body masses of unknown specimens are estimated by inserting the target individuals' values for the skeletal variable into the equation.

Numerous equations for estimating body mass can be found in the literature (Ruff et al. 1991; McHenry 1992; Aiello and Wood 1994; Ruff 1994; Grine et al. 1995; Kappelman 1996; Ruff et al. 1997, 2005; Auerbach and Ruff 2004, Spocter and Manger 2007). To date, three sets of equations have been derived for cranial variables (Aiello and Wood 1994; Kappelman 1996; Spocter and Manger 2007). Designed to address the paucity of associated postcranial remains in the hominin fossil record and the difficulty of associating such material to a given species, these equations have been used to generate body mass estimates for a variety of fossil hominin and primate specimens (Aiello and Wood 1994; Kappelman 1996; Wood and Collard 1996; Kordos and Begun 2001; Rightmire 2004; Spocter and Manger 2007; Churchill et al. 2012; Wu and Athreya 2013). More commonly, however, equations are based on postcranial variables. The majority of these equations employ measurements of the femur on the grounds that the lower limbs support the weight of the head, torso, and upper limbs and can therefore be expected to directly reflect body mass (Ruff et al. 1991). However, because factors like activity can also affect femoral morphology (e.g. Lieberman et al. 2004), and because mechanical loading may not be the same in past and present groups (Ruff 1994), equations based on overall body shape have also been developed (Ruff 1994; Ruff 2000a, b; Ruff et al. 2005). These equations employ skeletal variables that correspond to measures of body breadth (bi-iliac breadth) and stature (femoral length), and have been argued to be as good as the femur-based equations (Ruff et al. 1997; Auerbach and Ruff 2004). These equations have been used widely in palaeoanthropological, bioarchaeological, and forensic research (e.g. Ruff and Walker 1993; Arsuaga et al. 1999; Trinkaus and Jelinek 1997; Kurki et al. 2010; Myszka et al. 2012).

Despite their widespread use, the currently available equations may not be as reliable as they are usually assumed to be. Evidence for this comes from recent studies that obtained poor and/or inconsistent results when testing the equations against populations of known mass (Lorkiewicz-Muszyńska et al. 2013; Elliott et al. 2014; Elliott et al. 2015). In our exploration of the failure of many of the equations to yield reliable body mass estimates, it became apparent that one of the main sources of error is likely to be the data used to generate the equations (Elliott et al 2014; Elliott et al. 2015). Specifically, sample sizes have often been extremely small and the analyses have frequently employed indirect measures of key variables and/or unassociated mean data for the skeletal features and body mass (McHenry 1992; Aiello and Wood 1994; Kappelman 1996; Ruff 2000a; Ruff et al. 2005; Spocter and Manger 2007). As a result, the true nature of the relationship between the skeletal variables and body size has often not been adequately captured.

In light of the above, this study aimed to generate new equations for estimating body mass from cranial and postcranial skeletal features. By measuring variables directly from volume-rendered CT data on a large sample of modern humans of known body mass, our goal was to resolve some of the shortcomings of earlier equations and provide more reliable means for estimating mass than currently exist. In addition, by employing a known-mass test sample, drawn from the same population as the calibration group, we sought to evaluate the predictive competence of the resulting equations in a more robust way than has been possible in the past.

Materials and methods

Sample

This study used archived CT data from 128 male and 125 female deceased modern humans (range=18–90 years, male mean=48.1 years, female mean=51.2 years). CT scans were conducted at the Institute of Forensic Medicine (IFM) at the University of Zurich, Switzerland as part of routine forensic analyses (Thali et al. 2003), and the data were accessed through the IFM's secure server in accordance with their protocols. Query searches, record reviews, and visual inspection of the CT scans were used to select the sample. Exclusion criteria included individuals with skeletal abnormalities, trauma or prosthetics in the anatomical regions of interest, and any individual who was processed more than 3 days after death. Sex, age (years), body mass (kg), and stature (cm) were recorded for each individual. Population affinity is not recorded on post-mortem documentation in Switzerland. However, as more than 80 % of the Swiss population is of European descent (SFSO 2012), we considered the sample to be European. The full sample of 253 individuals was divided into two

Table 1 Summary statistics for calibration sample (n=203)

Variable	Females (n=100)			Males (n=103)			Combined sex (n=203)		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Weight (kg)	70.3	20.6	31.8-146	81.2	15.6	40.5–128.4	75.9	19.0	31.8–146
Stature (cm)	166.8	8.6	149.0–195.0	177.6	7.9	154.0–193.0	172.3	9.9	149.0–195.0
Age (years)	52	17	18-90	49	14.5	18–80	50.5	15.8	18–90

groups: a *calibration* sample (n=203) and a randomly chosen *test* sample (n=50), of roughly equal numbers of males and females each (Tables 1 and 2). The calibration sample was used to derive the regression equations, while the test sample was used to evaluate their accuracy and precision. Although other methods have been suggested (Smith 2002), such an external validation continues to be the most rigorous means of testing predictive equations (Harrell et al. 1996; Porter 1999; Giancristofaro and Salmaso 2007).

Imaging and 3D reconstruction protocols

CT scans were conducted at the IFM using a 128-slice, Siemens SOMATOM® Definition Flash, Dual-source CT scanner (Siemens Healthcare; Forchheim, Germany). Scans were taken at 120 kV, with a slice thickness of 0.75 mm (0.375 mm overlap), using bone convolution kernels (Thali et al. 2003). Milliampere-second (mAs) was automatically optimized using the Siemens CareDose® option. Scans were accessed from the IFM archive, and anatomical region volume rendered using OsiriX imaging software (64-bit extension, http://www.osirix-viewer.com) (Fig. 1). Skeletal elements were oriented in consistent planes and measured on the right side[1] to the nearest 0.1 mm using OsiriX tools. The accuracy of volume rendering skeletal models from CT has been demonstrated by several studies (Decker et al. 2011; Kim et al. 2012; Smyth et al. 2012). It was verified during this project by measuring, scanning, virtually reconstructing, and then virtually re-measuring an archaeological skull from the IFM's collection (Elliott et al. 2014). Measurement differences between the physical and virtual measurements in the test were less than 3 %, which was deemed an acceptable level of error.

[1] Due to the presence of fractures or prosthetics, 13 cases required the left femur to be measured. These measurements were included on the grounds that directional asymmetry in the lower limbs is small enough to be inconsequential for the purposes of estimating body mass (Auerbach and Ruff 2004; Ruff et al. 2012).

Skeletal variables

Twelve cranial and 24 postcranial variables were used for this study (Tables 3 and 4, Fig. 2). The cranial variables included six linear measurements selected on the basis of their performance in previous studies (Aiello and Wood 1994; Kappelman 1996; Spocter and Manger 2007). All of these studies identified orbital and foramen magnum areas as useful predictors of body mass in hominoid primates. However, as the variables were not measured in the same way in the studies in question, we included three calculations each for orbital and foramen magnum area: simple length-width (area=$L \times W$), in accordance with Aiello and Wood (1994) and Spocter and Manger (2007); area as an ellipse (area=$(\pi/4) \times L \times W$), following Spocter and Manger (2007); and a CAD-assisted method in which the perimeter of the feature was traced from a two-dimensional image, using a procedure similar to that of Kappelman (1996).

The postcranial variables included the two most commonly used measurements for predicting body mass from skeletal material: femoral head breadth (FHB) and bi-iliac breadth (BIB). FHB has been used in four sets of published regression equations (Ruff et al. 1991; McHenry 1992; Grine et al. 1995; Ruff et al. 2012) and is associated with the *mechanical* estimation approach (Ruff 2002; Auerbach and Ruff 2004). BIB[2] is used in conjunction with stature (STAT), in the second set of postcranial equations that are collectively referred to as the "morphometric" approach (Ruff 1994; Ruff et al. 1997, 2005). An additional 12 measurements of the femur were recorded in order to examine other claims that have been made in the literature. Femoral neck breadth (FNB) was measured in light of Ruff's (1991) suggestion that femoral neck size will exhibit a pattern of correlation with body mass between that of the head and shaft due to its intermediate location. Following Ruff and Hayes (1983) and Ruff (1991, 2000b), maximum femoral length (FLM) and measurements of medio-lateral shaft breadth (MLSB) and cortical breadth (MLCB), taken at multiple locations along the femur, were recorded to explore

[2] Because the sample was composed of recently deceased individuals with intact soft tissues, bi-iliac breadth was taken as *living breadth* and did not require the conversion from skeletal breadth as recommended by Ruff et al. (1997, 2005).

Table 2 Summary statistics for test sample (n=50)

Variable	Females (n=25)			Males (n=25)			Combined sex (n=50)		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Weight (kg)	66.3	12.8	48–93.7	83.0	19.8	52–142.3	74.6	18.5	48–142.3
Stature (cm)	164.0	6.5	152.0–180.0	176.9	7.9	156.0–190.0	170.5	9.7	152.0–190.0
Age (years)	48	14.3	29–80	44	11.7	23–70	46	13.0	23–88

how diaphyseal cross-sectional dimensions relate to weight. Lastly, shaft and cortical breadths were used to calculate 10 indices that correspond to cortical area (CA) and second moment of area (I) (Ruff et al. 1991). CA and I provide information about the cross-sectional geometry and strength of the femur (Ruff and Hayes 1983) and are calculated as $CA=\pi/4(D^2-d^2)$ and $I=\pi/64(D^4-d^4)$, respectively. These calculations were used here for comparison with Ruff and Hayes (1983) and Ruff et al. (1991). Supplementary tables 1–4 provide the summary data for the 36 variables in both samples.

Analyses

Using the calibration sample (n=203), correlation coefficients were calculated and regression equations derived for each variable for a combined-sex sample, as well as female-only (n=100) and male-only (n=103) subsamples. Regression equations were derived for each group using three methods: least square (LS), reduced major axis (RMA), and major axis (MA) regression because debate continues to surround the most appropriate choice for predicting an unknown quantity using regression. Konigsberg et al. (1998) evaluate five univariate estimation methods (*inverse calibration*, *classic calibration*, *major axis regression*, *reduced major axis regression*, and *ratio estimators*) and argue that least square regression (LSR) of body size (stature) on bone length (termed "inverse calibration") may be the most appropriate when the target specimen can be assumed to come from the same distribution as the calibration sample. Alternatively, they suggest that classic calibration—regression of bone length on body size, followed by solving for body size—may be better in situations where extrapolation is expected or uncertain. Despite this, most existing equations for estimating body mass have employed LSR and inverse calibration on the grounds that it minimizes the estimation error of the dependent variable (in this case, body mass) (Ruff et al. 1991). LSR also has the advantage of having correction factors available to account for biases inherent in the technique (Smith 1993). Alternative methods like RMA or MA have been argued to accept more uncertainty in the variables and produce better results when extrapolating beyond the range of the original dataset (Ruff et al. 1991; Auerbach and Ruff 2004). However, consensus on the most appropriate approach has not been reached and body mass continues to be estimated from LSR (Pomeroy and Stock 2012) and RMA (Ruff et al. 2012). Consequently, we derived equations with all three methods. To avoid problems associated with non-normal distributions, the data were logarithmically transformed (base 10). The standard error of the estimate (SEE) was calculated according to the formula $SEE=\sqrt{\sum(y-y')^2/n-2}$ (Hinton 2004). Two correction factors—the smearing estimate (SE) and ratio estimator (RE)[3]—were also calculated to account for de-transformation biases when converting from logarithmically transformed units into arithmetic units in LS regression (Smith 1993).

Once the regression equations were derived, individual skeletal measurements from the test sample were entered into the corresponding formula and an estimated body mass calculated in kilogrammes. For the LS regression equations, the estimated mass was corrected by multiplying it by the average of the SE and RE, to obtain a final estimated body mass (EBM) (Smith 1993). EBMs were then compared to the known masses for the individuals. Raw differences (RD) were calculated as known-mass EBM. Percentage differences were calculated as percent error using the formula (PE)=(known−EBM)/known*100 (Wu et al. 1995)[4]. Absolute percentage

Fig. 1 Example of a three-dimensionally reconstructed skeletal element, volume-rendered from CT data by OsiriX

[3] SE is calculated as $SE=1/n\sum\exp(\log r_i)$, while RE is calculated as RE=/, "where y is the observed value of the dependent variable Y for the ith observation on the original measurement scale and z, is the predicted value for the ith observation, detransformed back to the original measurement scale without correction" (Aiello and Wood, 1994:413).

[4] PEs provide the directional difference between the known and estimated masses. Positive PEs indicate that known mass>estimated mass (equation underestimates mass), and negative PEs indicate that known mass<estimated mass (equation overestimates mass).

Table 3 Cranial variables used in this study

	Variable	Description	Reference
1	BORB	Orbital breadth: distance between maxillofrontale and ectoconchion—in mm	Aiello and Wood 1994
2	HORB	Orbital height: distance between superior and inferior orbital margins, taken at a right angle to BORB—in mm	Aiello and Wood 1994
3	BIOR	Biorbital breadth: distance between two ectoconchion—in mm	Aiello and Wood 1994
4	BPOR	Biporionic breadth: distance from porion to porion—in mm	Aiello and Wood 1994
5	LFM	Foramen magnum length: distance between basion and opisthion—in mm	Aiello and Wood 1994
6	BFM	Foramen magnum breadth: distance in the coronal plane between the inner margins of the foramen magnum—in mm	Aiello and Wood 1994
7	ORBA1	Orbital area ($b \times h$): product of breadth × height—in mm^2	Aiello and Wood 1994
8	ORBA2	Orbital area (ellipse): calculated from breadth × height as an ellipse—in mm^2	Spocter and Manger 2007
9	ORBA3	Orbital area (CAD): calculated from perimeter margin using area function of ImageJ—in mm^2	Kappelman 1996
10	FMA1	Foramen magnum area ($b \times h$): product of breadth × height—in mm^2	Aiello and Wood 1994
11	FMA2	Foramen magnum area (ellipse): calculated from breadth × height as an ellipse—in mm^2	Spocter and Manger 2007
12	FMA3	Foramen magnum area (CAD): calculated from perimeter margin using area function of ImageJ—in mm^2	This study

differences (|PE|) were also calculated to assess the magnitude of the difference between the estimated and known masses (Aiello and Wood 1994; Ruff et al. 2005). Mean RDs, PEs, and |PE|s were calculated for males and females as well as the whole test sample. Differences were plotted and Wilcoxon signed-rank tests used to establish their significance. Lastly, in keeping with previous studies (e.g. Dagosto and Terranova 1992; Aiello and Wood 1994), we also calculated the percentage of individuals whose EBM fell within ±20 % of their known mass. All analyses were conducted in "R" (R Core Team 2010).

Evaluation criteria

To evaluate the predictive competence of a given equation, we used the same acceptance criteria as we employed in two recent tests of the validity of published cranial and postcranial equations (Elliott et al. 2014; Elliott et al. 2015). Specifically, equations were considered valid when absolute percent errors were below 19 % and at least 50 % of the individuals were estimated within ±20 % of their known mass. These values were selected in light of ongoing debate regarding what constitutes *acceptable* levels of error for predictive analyses. With respect to percent errors, Dagosto and Terranova (1992) considered 15–30 % to be inaccurate in interspecific analyses. In contrast, Spocter and Manger (2007) accepted prediction errors of 10–16 % for some of their analyses, while Aiello and Wood (1994) considered errors of 15–19 % to be acceptable. Regarding the number of estimates that should fall close to the actual mass, Ruff et al. (2005) suggest that a reliable intraspecific equation might be expected to estimate the majority of individuals within 10–15 % of their known mass. Others suggest that an equation can be considered reliable if it estimates 60–70 % of the specimens within 20 % of their known mass (Barrickman 2008). Due to the broader range of variation, interspecific analyses are expected to perform worse than intraspecific ones (Smith 2002). Consequently, interspecific studies often have even lower limits—accepting equations that estimate 50 % or more of the sample within ±20 % of known mass (e.g. Dagosto and Terranova 1992; Aiello and Wood 1994). Given the variability in acceptance criteria, we used the "19 % absolute error" and "50 % within 20 %" criteria as our limits of acceptance for consistency and because they provided a relatively lenient baseline to assess the equations' predictive competence.

Results

Regression equations

Tables 5 and 6 provide the LSR, MA, and RMA regression statistics for the cranial and postcranial variables, derived from the combined-sex calibration sample. Correlation coefficients for the cranial variables range from 0.06 ($p=0.36$) for orbital height (HORB) to 0.33 ($p=0.00$) for biporionic breadth (BPOR). For the postcranial variables, correlation coefficients range from −0.03 ($p=0.63$) for medio-lateral cortical breadth at 80 % of femoral shaft length (MLCB80) to 0.59 ($p=0.00$) for the index of cortical area at 50 % of femoral shaft length (CA50). SEEs for the cranial variables are consistent at 0.11, while SEEs vary slightly for the postcranial variables (0.09–0.11).

Tables 7 and 8 provide the cranial and postcranial regression equations derived from the female ($n=100$) calibration sample. The correlation coefficients for the cranial measurements vary from −0.01 for foramen magnum breadth (BFM) to 0.12 for BPOR. However, none of the correlations is

Table 4 Postcranial variables used in this study

	Variable	Description	Reference
13	FHB	Femoral head breadth: superior-inferior breadth perpendicular to the cervical axis—in mm	Ruff et al. 1991
14	BIB	Bi-iliac breadth: maximum pelvic breadth taken across the iliac crests, taken as "Living BIB"—in mm	Ruff 1991, 1994
15	FNB	Femoral neck breadth: minimum superior-inferior breadth at the point of deepest concavity of the superior surface—in mm	Ruff et al. 1991
16	FLM	Max femoral length: length of femur from the most superior point on the head to the most inferior point on the distal condyles—in mm	Buikstra and Ubelaker 1994
17	MLSB80	Medio-lateral shaft breadth at 80 % of total femur length, as measured from the distal end—in mm	Ruff and Hayes 1983
18	MLCB80	Sum of medial and lateral cortical breadths at 80 % of total femur length, as measured from the distal end—in mm	Ruff and Hayes 1983
19	CA80	Index of cortical area at 80 %: calculated from medio-lateral shaft (D) and cortical (d) breadth using $\pi/4(D^2-d^2)$—in mm^2	Ruff et al. 1991
20	I80	Index of second moment of area at 80 %: calculated from shaft (D) and cortical breadth (d) using $\pi/64(D^4-d^4)$—in mm^4	Ruff et al. 1991
21	MLSB65	Medio-lateral shaft breadth at 65 % of total femur length, as measured from the distal end—in mm	Ruff and Hayes 1983
22	MLCB65	Sum of medial and lateral cortical breadths at 65 % of total femur length, as measured from the distal end—in mm	Ruff and Hayes 1983
23	CA65	Index of cortical area at 65 %: calculated from shaft (D) and cortical (d) breadth using $\pi/4(D^2-d^2)$—in mm^2	Ruff et al. 1991
24	I65	Index of second moment of area at 65 %: calculated from shaft (D) and cortical breadth (d) using $\pi/64(D^4-d^4)$—in mm^4	Ruff et al. 1991
25	MLSB50	Medio-lateral shaft breadth at 50 % of total femur length, as measured from the distal end—in mm	Ruff and Hayes 1983
26	MLCB50	Sum of medial and lateral cortical breadths at 50 % of total femur length, as measured from the distal end—in mm	Ruff and Hayes 1983
27	CA50	Index of cortical area at 50 %: calculated from shaft (D) and cortical (d) breadth using $\pi/4(D^2-d^2)$—in mm^2	Ruff et al. 1991
28	I50	Index of second moment of area at 50 %: calculated from shaft (D) and cortical breadth (d) using $\pi/64(D^4-d^4)$—in mm^4	Ruff et al. 1991
29	MLSB35	Medio-lateral shaft breadth at 35 % of total femur length, as measured from the distal end—in mm	Ruff and Hayes 1983
30	MLCB35	Sum of medial and lateral cortical breadths at 35 % of total femur length, as measured from the distal end—in mm	Ruff and Hayes 1983
31	CA35	Index of cortical area at 35 %: calculated from shaft (D) and cortical (d) breadth using $\pi/4(D^2-d^2)$—in mm^2	Ruff et al. 1991
32	I35	Index of second moment of area at 35 %: calculated from shaft (D) and cortical breadth (d) using $\pi/64(D^4-d^4)$—in mm^4	Ruff et al. 1991
33	MLSB20	Medio-lateral shaft breadth at 20 % of total femur length, as measured from the distal end—in mm	Ruff and Hayes 1983
34	MLCB20	Sum of medial and lateral cortical breadths at 20 % of total femur length, as measured from the distal end—in mm	Ruff and Hayes 1983
35	CA20	Index of cortical area at 20 %: calculated from shaft (D) and cortical (d) breadth using $\pi/4(D^2-d^2)$—in mm^2	Ruff et al. 1991
36	I20	Index of second moment of area at 20 %: calculated from shaft (D) and cortical breadth (d) using $\pi/64(D^4-d^4)$—in mm^4	Ruff et al. 1991

significant at $\propto=0.05$. Correlation coefficients for the postcranial measurements vary from −0.11 ($p=0.28$) for medio-lateral cortical breadth at 35 % (MLCB35) to 0.59 ($p=0.00$) for the index of CA50. SEEs for the cranial variables are again consistent, this time at 0.12. SEEs for the postcranial variables range from 0.10 to 0.12.

In the male sample, the correlation coefficients for the cranial measurements (Table 9) vary from 0.05 ($p=0.62$) for HORB to 0.28 ($p=0.00$) for BPOR. Correlations for the postcranial variables (Table 10) range from −0.02 ($p=0.81$) for medio-lateral cortical breadth at 35 % of femoral shaft (MLCB35) to 0.42 ($p=0.00$) for the index of CA50. SEEs for the cranial variables are consistent at 0.09 and range from 0.08 to 0.09 for the postcranial variables.

Prediction accuracy

Tables 11, 12, 13, 14, 15, and 16 summarize the directional and absolute differences, as well as the percentage of individuals whose body mass was estimated within ±20 % of their true mass, for each sample using the LSR equations. As the major axis (MA) and reduced major axis (RMA) regression equations produced significantly higher rates of error for most skeletal measurements in all three samples, their results are provided in the supplementary materials (Suppl. Tables 5–10) and only the least square (LS) regression results are discussed here.

In the combined-sex test sample, the cranial variables with the best predictive accuracy based on the absolute percentage

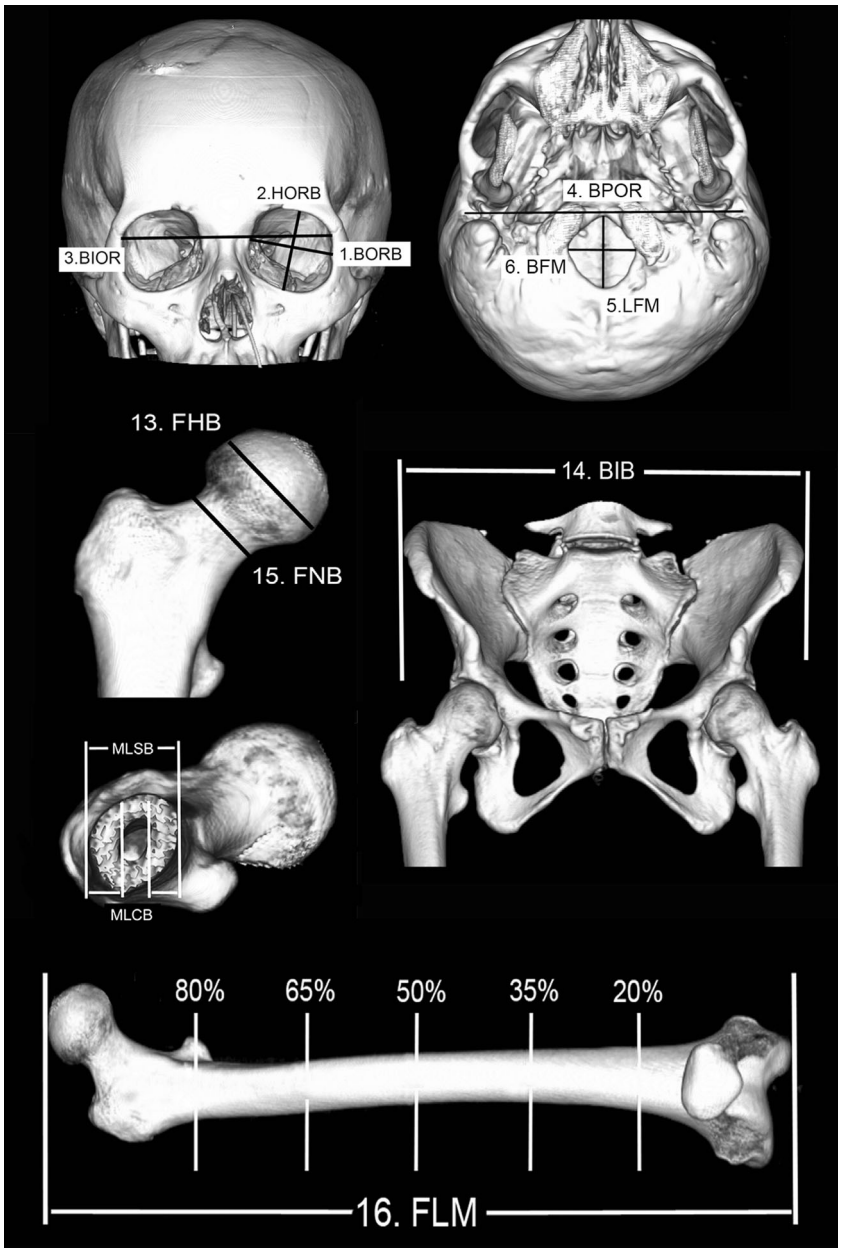

Fig. 2 Cranial and post-cranial measurement variables and cross-sectional locations used in this study

errors and number of individuals estimated within ±20 % of known mass were BPOR and biorbital breadth (BIOR) (Table 11). These two variables resulted in mean absolute percent errors (|PE|) of 17.1 and 18.1 %, respectively. Both variables estimated 64 % of the body masses in the test sample within ±20 % of the known mass. The worst performing cranial variable was foramen magnum breadth, which produced a |PE| of 19.6 % and estimated 56 % of the sample masses within ±20 % of known mass.

The postcranial variable with the best predictive accuracy in the combined-sex test sample was the index of cortical area at 35 % of femoral length (CA35)—returning a |PE| of 11.9 % and estimating 82 % of the sample within ±20 % of known mass (Table 16). The five indices of second moment of area (*I*) returned the lowest accuracy rates, resulting in |PE|s exceeding 100 % and estimating no individuals within ±20 % of known mass.

The best performing cranial variables in the female test sample were BIOR and orbital area calculated as $b \times h$ (ORBA1) (Table 13). These variables resulted in mean |PE|s of 17.8 and 19.2 %, respectively, and estimated 56 % of the body masses in the sample within ±20 % of known mass. The least accurate cranial variable was orbital area, calculated as an ellipse (ORBA3). The equation for this variable returned a |PE| of 18.3 % and estimated 48 % of the sample masses within ±20 % of known mass.

The best postcranial predictor of body mass in females was the index of cortical area at 80 % of femur length (CA80),

Table 5 Least squares (LS), major axis (MA), and reduced major axis (RMA) regression equations for cranial variables. Combined-sex calibration sample (n=203)

Variable	r	p value	LSR regression				MA regression		RMA regression	
			Slope	Intercept	SEE	CF	Slope	Intercept	Slope	Intercept
BORB	0.21	0.003	0.951	0.385	0.11	1.03	21.47	−31.59	4.62	−5.34
HORB	0.06	0.362	0.245	1.491	0.11	1.03	55.24	−82.89	3.81	−3.98
BIOR	0.30	0.000	1.566	−1.236	0.11	1.03	16.67	−31.18	5.20	−8.43
BPOR	0.33	0.000	1.527	−1.285	0.11	1.03	13.23	−25.44	4.59	−7.61
LFM	0.21	0.012	0.730	0.746	0.11	1.03	15.59	−22.06	3.51	−3.52
BFM	0.13	0.063	0.384	1.305	0.11	1.03	19.92	−27.27	2.94	−2.43
ORBA1	0.17	0.013	0.492	0.345	0.11	1.03	14.34	−42.47	2.83	−6.89
ORBA2	0.17	0.013	0.492	0.397	0.11	1.03	14.34	−40.96	2.83	−6.59
ORBA3	0.18	0.012	0.522	0.303	0.11	1.04	14.81	−42.51	2.95	−6.97
FMA1	0.19	0.007	0.341	0.844	0.11	1.03	6.86	−18.69	1.81	−3.57
FMA2	0.18	0.007	0.341	0.880	0.11	1.03	6.86	−17.97	1.81	−3.38
FMA3	0.20	0.004	0.370	0.814	0.11	1.03	6.50	−16.65	1.83	−3.34

All data log10 transformed

SEE standard error of estimate, *CF* correction factor (mean of SE and RE)

Table 6 Least squares (LS), major axis (MA), and reduced major axis (RMA) regression equations for postcranial variables. Combined-sex calibration sample (n=203)

Variable	r	p value	LSR regression				MA regression		RMA regression	
			Slope	Intercept	SEE	CF	Slope	Intercept	Slope	Intercept
BIB	0.36	0.000	1.43	−1.65	0.10	1.03	10.6	−24.1	4.0	−7.96
FLM	0.33	0.000	1.28	−1.54	0.10	1.03	11.3	−28.1	3.9	−8.54
FHB	0.40	0.000	1.33	−0.38	0.10	1.03	7.6	−11.0	3.3	−3.72
FNB	0.43	0.000	1.12	0.16	0.10	1.03	5.3	−6.3	2.6	−2.11
MLSB80	0.41	0.000	1.25	−0.06	0.10	1.03	6.6	−8.4	3.0	−2.78
MLCB80	−0.03	0.630	− 0.05	1.93	0.11	1.03	−23.3	30.3	−1.5	3.7
CA80	0.52	0.000	0.71	−0.18	0.09	1.03	1.8	−3.3	1.4	−2.1
I80	0.44	0.000	0.33	0.24	0.10	1.03	0.54	−0.77	0.75	−1.8
MLSB65	0.49	0.000	1.35	−0.14	0.10	1.03	5.1	−5.6	2.7	−2.2
MLCB65	0.14	0.039	− 0.15	2.00	0.11	1.03	−1.4	3.1	−1.0	2.8
CA65	0.55	0.000	0.73	− 0.18	0.09	1.03	1.6	−2.7	1.3	−1.9
I65	0.50	0.000	0.34	0.28	0.10	1.03	0.5	−0.4	0.7	−1.3
MLSB50	0.52	0.000	1.58	−0.47	0.09	1.03	5.3	−6.0	3.0	−2.6
MLCB50	0.15	0.038	−0.17	2.02	0.11	1.03	−2.3	4.1	−1.1	2.96
CA50	0.59	0.000	0.85	−0.53	0.09	1.03	1.8	−3.2	1.4	−2.18
I50	0.53	0.000	0.40	0.03	0.09	1.03	0.6	−0.9	0.8	−1.60
MLSB35	0.44	0.000	1.29	−0.08	0.10	1.03	6.1	−7.4	2.9	−2.58
MLCB35	0.00	0.992	0.00	1.87	0.11	1.03	0.0	−844.9	0.0	0.33
CA35	0.56	0.000	0.82	−0.44	0.09	1.03	1.9	−3.5	1.5	−2.23
I35	0.47	0.000	0.35	0.21	0.97	1.03	0.6	−0.7	0.7	−1.65
MLSB20	0.32	0.000	0.89	0.43	0.10	1.03	7.8	−10.8	2.8	−2.66
MLCB20	0.04	0.597	0.06	1.79	0.11	1.03	22.3	−30.0	1.5	−0.28
CA20	0.50	0.000	0.62	0.10	0.10	1.03	1.5	−2.5	1.2	−1.68
I20	0.42	0.000	0.31	0.32	0.10	1.03	0.5	−0.7	0.7	−1.84

All data log10 transformed

SEE standard error of estimate, *CF* correction factor (mean of SE and RE)

Table 7 Least squares (LS), major axis (MA), and reduced major axis (RMA) regression equations for cranial variables. Female calibration sample (n=100)

Variable	r	p value	LSR regression				MA regression		RMA regression	
			Slope	Intercept	SEE	CF	Slope	Intercept	Slope	Intercept
BORB	0.04	0.721	0.22	1.50	0.12	1.05	160.6	−246.5	6.0	−7.4
HORB	0.08	0.448	0.33	1.33	0.12	1.04	52.1	−78.1	4.2	−4.7
BIOR	0.15	0.127	1.00	−0.14	0.12	1.04	41.3	−79.6	6.5	−11.0
BPOR	0.12	0.247	0.71	0.38	0.12	1.04	50.5	−101.6	6.1	−10.6
LFM	0.07	0.500	0.30	1.38	0.12	1.03	60.4	−90.3	4.3	−4.8
BFM	−0.01	0.914	−0.04	1.88	0.12	1.05	−280.4	409.3	−3.4	6.7
ORBA1	0.08	0.446	0.24	1.07	0.12	1.07	37.3	−113.1	3.2	−8.0
ORBA2	0.08	0.446	0.24	1.10	0.12	1.06	37.3	−109.2	3.2	−7.6
ORBA3	0.06	0.560	0.19	1.25	0.12	1.04	50.3	−148.5	3.3	−8.0
FMA1	0.03	0.790	0.06	1.66	0.12	1.03	63.3	−186.7	2.2	−4.6
FMA2	0.03	0.790	0.06	1.66	0.12	1.04	63.3	−180.1	2.2	−4.4
FMA3	0.03	0.754	0.07	1.64	0.12	1.04	52.0	−145.2	2.1	−4.2

All data log10 transformed

SEE standard error of estimate, *CF* correction factor (mean of SE and RE)

which returned a |PE| of 13.8 % and estimated 80 % of the sample within ±20 % of known mass (Table 18). Mediolateral cortical breadth at 35 % of femur length, as well as the five second moment of area indices, all returned |PE|s in excess of 100 % and failed to estimate any individuals within ±20 % of known mass.

For the male sample, the best cranial predictor of body mass was BPOR (Table 15). This variable returned a |PE| of 15.9 % and estimated 76 % of the sample within ±20 % of known mass. The least accurate predictors were HORB, BFM, orbital area as an ellipse (ORBA2), and ORBA1. These variables resulted in |PE|s between 17.6 and 19.2 % and estimated 60 % of the sample within ±20 % of known mass.

With regard to the postcranial variables (Table 16), body mass was estimated best in the male sample using mediolateral shaft breadth at 65 % of femur length (MLSB65). The LSR equation for this measurement resulted in a |PE| of 17.6 % and estimated 65 % of the male sample's body masses within ±20 % of known mass. As in the other two test samples, the indices of second moment of area performed particularly poorly—with four of the five returning |PE|s in excess of 66 % and estimating only 4 % of the body masses in the male sample within ±20 % of known mass.

Discussion

The goal of the study reported here was to improve the estimation of body mass from skeletal remains by deriving new regression equations based on more robust data than had been available to previous studies. To achieve this, we used a large calibration sample consisting of both males and females; employed skeletal elements that were complete and undistorted; and derived regression equations from directly measured skeletal variables matched to individual, associated body masses. The resulting equations were then evaluated against a known-mass test sample, using acceptance criteria derived from the literature.

The results show that 6 of the 12 cranial equations can be considered valid in the combined-sex sample. Fourteen of the 24 postcranial equations also met the criteria for acceptance in this group. In the female-only sample, all but 1 cranial equation (ORBA3) and 14 of 24 postcranial equations were valid. In the male-only sample, all the cranial equations met the acceptance criteria, while 11 of the postcranial equations were acceptable predictors. Thus, the majority of the new equations can be considered reliable estimators of mass, according to the assessment criteria.

Table 17 summarizes a comparison between the results of our equations and those of previous studies as tested against our combined-sex test sample (n=50) (Supplementary Table 11 provides the comparisons for the female and male samples). The test employed the three *best* cranial predictors from Aiello and Wood (1994) (orbital area, orbital height, and biporionic breadth) and the femoral head breadth equations provided by Ruff et al. (1991), McHenry (1992), Grine et al. (1995), and Ruff et al. (2012). Aiello and Wood's (1994) equations were used rather than those of Kappelman (1996) and Spocter and Manger (2007) because they produced lower

Table 8 Least squares (LS), major axis (MA), and reduced major axis (RMA) regression equations for postcranial variables. Female calibration sample (n=100)

Variable	r	p value	LSR regression			MA regression			RMA regression	
			Slope	Intercept	SEE	CF	Slope	Intercept	Slope	Intercept
BIB	0.32	0.001	1.27	−1.27	0.11	1.03	11.8	−26.9	4.0	−7.9
FLM	0.14	0.156	0.68	0.04	0.12	1.03	31.6	−81.8	4.7	−10.7
FHB	0.15	0.128	0.82	0.47	0.11	1.04	33.8	−54.3	5.4	−7.1
FNB	0.27	0.006	1.05	0.25	0.12	1.04	13.3	−18.2	3.9	−4.0
MLSB80	0.33	0.001	1.36	−0.25	0.11	1.04	12.2	−16.8	4.2	−4.6
MLCB80	−0.25	0.013	−0.44	2.36	0.12	1.04	−5.1	8.0	−1.8	4.0
CA80	0.53	0.000	0.99	−0.98	0.10	1.04	2.8	−6.2	1.9	−3.4
I80	0.38	0.000	0.40	−0.09	0.11	1.05	1.1	−3.6	1.1	−3.2
MLSB65	0.43	0.000	1.43	−0.27	0.11	1.05	7.3	−8.8	3.4	−3.1
MLCB65	−0.36	0.000	−0.40	2.19	0.11	1.05	−1.4	3.1	−1.1	2.8
CA65	0.52	0.000	0.83	−0.47	0.10	1.03	2.3	−4.4	1.6	−2.6
I65	0.44	0.000	0.37	0.16	0.11	1.03	0.7	−1.2	0.8	−2.0
MLSB50	0.49	0.000	1.82	−0.83	0.11	1.05	7.3	−8.8	3.7	−3.6
MLCB50	−0.30	0.002	−0.37	2.18	0.11	1.03	−1.8	3.6	−1.2	3.0
CA50	0.59	0.000	1.02	−0.98	0.10	1.02	2.4	−4.8	1.7	−3.0
I50	0.50	0.000	0.47	−0.27	0.11	1.01	0.9	−2.1	0.9	−2.4
MLSB35	0.37	0.000	1.39	−0.24	0.11	1.04	9.4	−12.1	3.7	−3.7
MLCB35	−0.11	0.279	−0.15	2.00	0.12	1.05	−6.0	8.8	−1.4	3.4
CA35	0.57	0.000	1.09	−1.18	0.10	1.04	2.8	−5.8	1.9	−3.4
I35	0.42	0.000	0.40	−0.03	0.11	1.05	0.9	−2.3	1.0	−2.6
MLSB20	0.21	0.034	0.72	0.68	0.12	1.04	14.7	−21.6	3.4	−3.6
MLCB20	−0.12	0.243	−0.20	2.11	0.12	1.05	−9.3	15.0	−1.7	4.2
CA20	0.54	0.000	0.82	−0.50	0.10	1.05	2.1	−4.1	1.5	−2.5
I20	0.36	0.000	0.33	0.17	0.11	1.06	0.8	−2.1	0.9	−2.7

All data log10 transformed

SEE standard error of estimate, *CF* correction factor (mean of SE and RE)

Table 9 Least squares (LS), major axis (MA), and reduced major axis (RMA) regression equations for cranial variables. Male calibration sample (n=103)

Variable	r	p value	LSR regression				MA regression		RMA regression	
			Slope	Intercept	SEE	CF	Slope	Intercept	Slope	Intercept
BORB	0.09	0.384	0.35	1.35	0.09	1.02	43.3	−66.1	4.0	−4.4
HORB	0.05	0.620	0.14	1.67	0.09	1.04	52.3	−78.4	2.9	−2.6
BIOR	0.19	0.051	0.92	0.07	0.09	1.02	23.5	−45.0	4.7	−7.6
BPOR	0.28	0.004	1.30	−0.79	0.09	1.01	15.9	−31.1	4.6	−7.7
LFM	0.20	0.039	0.54	1.06	0.09	1.03	11.4	−15.7	2.7	−2.2
BFM	0.15	0.134	0.35	1.39	0.09	1.02	12.9	−17.0	2.3	−1.5
ORBA1	0.09	0.359	0.22	1.23	0.09	1.01	21.4	−64.5	2.4	−5.5
ORBA2	0.09	0.359	0.22	1.25	0.09	1.01	21.4	−62.2	2.4	−5.2
ORBA3	0.12	0.237	0.30	1.00	0.09	1.02	18.0	−52.3	2.5	−5.7
FMA1	0.20	0.043	0.28	1.04	0.09	1.04	3.9	−9.8	1.4	−2.4
FMA2	0.20	0.043	0.28	1.07	0.09	1.04	3.9	−9.4	1.4	−2.2
FMA3	0.21	0.035	0.32	0.99	0.09	1.02	4.3	−10.5	1.5	−2.4

All data log10 transformed

SEE standard error of estimate, *CF* correction factor (mean of SE and RE)

Table 10 Least squares (LS), major axis (MA), and reduced major axis (RMA) regression equations for postcranial variables. Male calibration sample (n=103)

Variable	r	p value	LSR regression				MA regression		RMA regression	
			Slope	Intercept	SEE	CF	Slope	Intercept	Slope	Intercept
BIB	0.34	0.000	1.23	−1.11	0.08	1.01	10.0	−22.7	3.6	−7.0
FLM	0.29	0.003	1.07	−0.97	0.09	1.03	11.5	−28.9	3.6	−7.9
FHB	0.40	0.000	1.57	−0.78	0.08	1.02	9.3	−14.0	3.9	−4.8
FNB	0.34	0.000	1.04	0.28	0.08	1.02	7.9	−10.5	3.0	−2.8
MLSB80	0.26	0.008	0.74	0.74	0.09	1.02	9.6	−13.2	2.8	−2.6
MLCB80	0.06	0.517	0.07	1.81	0.09	1.03	2.9	−1.7	1.1	0.6
CA80	0.29	0.003	0.39	0.75	0.09	1.03	2.5	−5.3	1.3	−2.0
I80	0.28	0.005	0.20	0.92	0.09	1.01	0.4	0.1	0.7	−1.7
MLSB65	0.34	0.000	1.00	0.41	0.09	1.00	7.7	−9.8	2.9	−2.5
MLCB65	−0.08	0.406	−0.07	1.97	0.09	1.02	−0.3	2.2	−0.9	2.8
CA65	0.39	0.000	0.53	0.38	0.08	1.02	2.1	−4.2	1.4	−2.0
I65	0.35	0.000	0.25	0.71	0.08	1.03	0.4	−0.2	0.7	−1.5
MLSB50	0.35	0.000	1.12	0.22	0.08	1.02	8.3	−10.6	3.2	−2.9
MLCB50	−0.15	0.133	−0.15	2.04	0.09	1.03	−0.8	2.7	−1.0	2.9
CA50	0.42	0.000	0.64	0.07	0.08	1.03	2.5	−5.1	1.5	−2.4
I50	0.36	0.000	0.29	0.56	0.08	1.00	0.6	−0.7	0.8	−1.8
MLSB35	0.24	0.014	0.76	0.74	0.09	1.02	11.7	−16.0	3.1	−2.9
MLCB35	−0.02	0.814	−0.03	1.93	0.09	1.03	−9.3	13.0	−1.1	3.2
CA35	0.36	0.000	0.58	0.26	0.08	1.01	3.1	−6.8	1.6	−2.7
I35	0.27	0.005	0.22	0.84	0.09	1.03	0.5	−0.5	0.8	−2.0
MLSB20	0.20	0.048	0.48	1.11	0.09	1.03	10.7	−15.6	2.5	−2.1
MLCB20	0.13	0.196	0.15	1.69	0.09	1.02	2.7	−2.1	1.2	0.2
CA20	0.23	0.017	0.27	1.10	0.09	1.03	1.8	−3.4	1.2	−1.5
I20	0.24	0.014	0.16	1.06	0.09	1.05	0.3	0.5	0.7	−1.6

All data log10 transformed

SEE standard error of estimate, *CF* correction factor (mean of SE and RE)

Table 11 Summary of difference between known and estimated mass using LSR equations for cranial variables. Combined-sex test sample (n=50), ordered by percentage of individuals estimated within 20 % of known mass

Variable	Directional difference		Absolute difference		20 % (%)	t test p value
	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)		
BPOR	−0.51 (16.2)	−5.38 (21.2)	12.28 (10.4)	17.13 (13.5)	64.0	0.82
BIOR	−1.26 (17.1)	−6.78 (22.9)	12.83 (11.3)	18.11 (15.3)	64.0	0.6
FMA3	−1.16 (17.8)	−6.79 (22.9)	13.39 (11.6)	18.91 (14.3)	58.3	0.65
BORB	−1.10 (17.8)	−6.83 (23.6)	13.77 (11.2)	19.33 (14.9)	58.0	0.66
LFM	−0.18 (18.2)	−5.49 (22.8)	13.65 (11.8)	18.78 (13.7)	58.0	0.94
ORBA1	−0.85 (18.0)	−6.54 (23.6)	13.71 (11.6)	19.18 (15.0)	58.0	0.74
ORBA2	−0.92 (18.0)	−6.63 (23.6)	13.71 (11.6)	19.20 (15.0)	58.0	0.72
FMA1	−0.43 (17.8)	−5.85 (22.9)	13.57 (11.4)	18.87 (14.0)	58.0	0.87
FMA2	−0.47 (17.8)	−5.90 (22.9)	13.57 (11.4)	18.88 (14.0)	58.0	0.85
ORBA3	−2.42 (18.0)	−8.67 (23.8)	13.58 (11.8)	19.47 (15.9)	56.2	0.36
HORB	−1.08 (18.4)	−7.08 (24.2)	14.17 (11.7)	19.90 (15.3)	56.0	0.68
BFM	−0.94 (18.0)	−6.72 (23.7)	13.95 (11.2)	19.61 (14.7)	56.0	0.71

Table 12 Summary of difference between known and estimated mass using LSR equations for postcranial variables. Combined-sex test sample (n= 50), ordered by percentage of individuals estimated within 20 % of known mass

Variable	Directional difference		Absolute difference		20 % (%)	t test p value
	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)		
CA35	0.8 (11.3)	−1.6 (14.6)	8.9 (6.8)	11.9 (8.4)	82.0	0.62
CA65	−2.1 (12.5)	−5.9 (16.2)	10.2 (7.3)	14.1 (9.9)	80.0	0.24
MLSB50	1.0 (12.6)	−1.7 (16.0)	9.2 (8.6)	12.3 (10.2)	80.0	0.58
CA50	2.3 (11.7)	0.7 (14.4)	8.6 (8.2)	11.1 (8.9)	80.0	0.17
MLSB65	1.4 (13.4)	−1.4 (16.5)	9.9 (9.1)	13.0 (10.1)	78.0	0.46
CA80	1.1 (12.9)	−1.7 (16.5)	9.6 (8.5)	12.8 (10.3)	76.0	0.56
MLSB35	0.1 (13.9)	−3.5 (18.1)	10.9 (8.5)	14.7 (10.8)	74.0	0.95
FHB	1.3 (15.7)	−2.3 (19.4)	12.1 (9.8)	16.2 (10.7)	70.0	0.55
CA20	−2.8 (14.2)	−7.5 (19.4)	11.7 (8.3)	16.5 (12.4)	70.0	0.17
MLSB80	−1.2 (14.6)	−5.5 (18.8)	11.1 (9.4)	15.4 (12.0)	68.0	0.56
BIB	3.0 (16.8)	−0.5 (21.4)	12.7 (11.3)	16.8 (13.0)	64.0	0.21
MLSB20	−0.1 (16.1)	−4.6 (20.9)	12.6 (9.9)	17.2 (12.5)	62.0	0.97
FLM	−0.4 (17.0)	−5.2 (21.6)	13.2 (10.6)	18.1 (12.5)	60.0	0.88
MLCB65	−1.0 (18.4)	−7.0 (24.8)	14.4 (11.3)	20.4 (15.5)	58.0	0.69
MLCB50	−0.3 (18.8)	−6.0 (24.8)	14.5 (11.7)	20.2 (15.4)	58.0	0.92
FNB	−2.2 (16.1)	−7.2 (21.1)	13.2 (9.3)	18.4 (12.4)	56.0	0.35
MLCB80	−1.6 (18.5)	−7.9 (26.2)	14.6 (11.4)	20.6 (18.3)	54.0	0.54
MLCB35	−1.7 (18.5)	−8.1 (24.7)	14.6 (11.4)	20.6 (15.6)	54.0	0.51
MLCB20	−2.6 (18.4)	−9.3 (24.8)	14.7 (11.2)	21.0 (15.9)	54.0	0.31
I80	−101.1 (16.0)	−144.1 (42.3)	101.1 (16.0)	144.1 (42.3)	0.0	0
I65	−108.4 (16.3)	−153.6 (41.0)	108.4 (16.3)	153.6 (41.0)	0.0	0
I50	−150.7 (23.4)	−211.0 (48.4)	150.7 (23.4)	211.0 (48.4)	0.0	0
I35	−117.4 (17.4)	−166.1 (44.6)	117.4 (17.4)	166.1 (44.6)	0.0	0
I20	−109.7 (20.7)	−156.6 (49.4)	109.7 (20.7)	156.6 (49.4)	0.0	0

Table 13 Summary of difference between known and estimated mass using LSR equations for cranial variables. Female test sample (n=25), ordered by percentage of individuals estimated within 20 % of known mass

Variable	Directional difference		Absolute difference		20 % (%)	t test p value
	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)		
BIOR	−4.4 (12.3)	−10.3 (20.3)	10.7 (7.2)	17.8 (13.8)	56.0	0.09
ORBA1	−5.8 (12.5)	−12.5 (20.8)	11.4 (7.4)	19.2 (14.6)	56.0	0.03
ORBA2	−5.1 (12.5)	−11.5 (20.6)	11.3 (7.0)	18.8 (13.9)	56.0	0.05
BORB	−5.1 (12.8)	−11.5 (20.9)	11.5 (7.3)	19.0 (14.1)	52.0	0.06
HORB	−3.7 (12.4)	−9.2 (20.1)	11.2 (6.1)	18.3 (12.0)	52.0	0.15
BPOR	−4.0 (12.7)	−9.7 (20.3)	10.9 (7.3)	17.9 (13.4)	52.0	0.13
LFM	−2.9 (12.5)	−8.0(20.0)	11.1 (6.2)	17.9 (11.5)	52.0	0.26
BFM	−4.8 (12.8)	−11.1 (20.9)	11.5 (7.1)	19.0 (13.7)	52.0	0.07
FMA1	−3.7 (12.8)	−9.3 (20.6)	11.3 (6.6)	18.5 (12.5)	52.0	0.17
FMA2	−3.6 (12.8)	−9.3 (20.6)	11.3 (6.6)	18.5 (12.5)	52.0	0.17
FMA3	−4.0 (12.8)	−9.8 (20.7)	11.4 (6.7)	18.7 (12.9)	52.0	0.13
ORBA3	−4.2 (12.5)	−10.1 (20.4)	11.2 (6.7)	18.3 (13.1)	48.0	0.1

Table 14 Summary of difference between known and estimated mass using LSR equations for postcranial variables. Female test sample (n=25), ordered by percentage of individuals estimated within 20 % of known mass

Variable	Directional difference		Absolute difference		20 % (%)	t test p value
	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)		
CA80	−3.6 (10.1)	−7.5 (16.4)	8.6 (6.2)	13.8 (11.3)	80.0	0.09
CA50	−3.6 (10.9)	−7.3 (16.8)	10.0 (5.3)	15.4 (9.4)	76.0	0.11
CA35	−3.0 (11.0)	−6.3 (17.2)	8.9 (6.9)	13.8 (11.7)	76.0	0.18
MLSB80	−2.7 (11.5)	−7.0 (18.4)	9.6 (6.7)	15.4 (11.9)	72.0	0.26
MLSB65	−2.6 (10.3)	−6.4 (16.4)	8.5 (6.1)	13.7 (10.8)	72.0	0.21
CA65	−4.8 (10.0)	−9.3 (16.0)	9.7 (5.3)	15.3 (10.0)	72.0	0.02
MLSB50	−2.3 (11.2)	−5.8 (17.8)	9.4 (6.3)	14.8 (11.2)	72.0	0.31
CA20	−3.2 (12.3)	−7.2 (19.7)	9.2 (8.6)	14.7 (14.8)	72.0	0.2
MLSB35	−3.0 (12.2)	−7.3 (18.9)	10.1 (7.2)	15.9 (12.1)	68.0	0.24
MLCB80	−5.4 (12.7)	−11.9 (21.0)	11.5 (7.5)	19.0 (14.6)	64.0	0.04
MLSB20	−3.4 (12.6)	−8.5 (19.7)	10.8 (7.0)	17.3 (12.3)	64.0	0.19
MLCB20	−4.5 (12.8)	−10.6 (21.2)	11.4 (7.0)	18.9 (13.9)	64.0	0.09
MLCB50	−4.4 (14.2)	−10.2 (23.7)	11.5 (9.2)	18.8 (17.3)	60.0	0.13
FHB	−3.3 (13.0)	−8.7 (20.8)	11.3 (6.9)	18.3 (12.6)	56.0	0.22
FNB	−3.1 (14.0)	−8.5 (22.1)	12.0 (7.5)	19.2 (13.4)	56.0	0.28
BIB	−2.4 (13.6)	−7.3 (21.5)	11.3 (7.6)	18.1 (13.2)	52.0	0.38
FLM	−3.8 (13.5)	−9.7 (21.2)	11.9 (7.0)	19.3 (12.6)	52.0	0.17
MLCB65	−7.0 (13.3)	−14.1 (22.5)	12.2 (8.5)	20.5 (16.6)	52.0	0.01
I80	−141.6 (19.5)	−221.9 (54.5)	141.6 (19.5)	221.9 (54.5)	0.0	0
I65	−130.1 (19.2)	−203.0 (46.6)	130.1 (19.2)	203.0 (46.6)	0.0	0
I50	−194.0 (32.8)	−301.3 (67.3)	194.0 (32.8)	301.3 (67.3)	0.0	0
MLCB35	−89.3 (14.4)	−143.3 (47.0)	89.3 (14.4)	143.3 (47.0)	0.0	0
I35	−140.4 (22.5)	−219.8 (55.6)	140.4 (22.5)	219.8 (55.6)	0.0	0
I20	−104. 1 (20.5)	−164.4 (50.1)	104.1 (20.5)	164.4 (50.1)	0.0	0

Table 15 Summary of difference between known and estimated mass using LSR equations for cranial equations. Male test sample (n=25), ordered by percentage of individuals estimated within 20 % of known mass

Variable	Directional difference		Absolute difference		20 % (%)	t test p value
	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)		
BPOR	3.7 (17.9)	0.1 (20.1)	13.4 (12.1)	15.9 (11.8)	76	0.3
BIOR	2.1 (19.5)	−2.5 (23.0)	14.5 (12.9)	17.7 (14.4)	72	0.6
ORBA3	0.6 (20.4)	−4.6 (23.7)	15.1 (13.3)	18.8 (14.7)	65.2	0.9
FMA3	0.8 (20.3)	−4.2 (22.4)	14.6 (13.8)	17.8 (13.7)	65.2	0.9
FMA2	0.1 (19.8)	−4.9 (22.7)	14.8 (12.7)	18.3 (13.8)	64	1.0
FMA1	0.2 (19.8)	−4.8 (22.7)	14.8 (12.7)	18.3 (13.8)	64	1.0
LFM	1.2 (20.5)	−3.7 (23.2)	15.4 (13.3)	18.7 (13.7)	64	0.8
BORB	2.1 (19.6)	−2.5 (22.9)	14.9 (12.6)	18.2 (13.6)	64	0.6
HORB	−0.2 (20.0)	−5.5 (24.0)	15.3 (12.4)	19.2 (14.9)	60	1.0
BFM	1.5 (19.3)	−3.1 (22.2)	14.4 (12.6)	17.6 (13.5)	60	0.7
ORBA2	2.6 (19.9)	−1.9 (23.1)	15.2 (12.8)	18.5 (13.5)	60	0.5
ORBA1	2.7 (19.9)	−1.8 (23.1)	15.2 (12.8)	18.5 (13.5)	60	0.5

Table 16 Summary of difference between known and estimated mass using LSR equations for postcranial equations. Male test sample (n=25), ordered by percentage of individuals estimated within 20 % of known mass

Variable	Directional difference		Absolute difference		20 % (%)	t test p value
	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)		
MLSB65	9.0 (19.6)	6.6 (20.3)	15.8 (14.4)	17.6 (11.7)	65	0.0
MLCB20	1.7 (20.5)	−3.2 (24.5)	15.7 (12.9)	19.4 (14.9)	64	0.7
MLSB35	7.8 (21.1)	4.4 (23.7)	16.9 (14.6)	19.6 (13.5)	64	0.1
CA80	9.2 (19.9)	6.6 (21.2)	16.4 (14.4)	18.4 (11.9)	64	0.0
MLCB65	0.2 (19.6)	−4.9 (23.5)	15.3 (11.9)	19.1 (14.1)	60	1.0
BIB	5.2 (19.9)	1.5 (23.0)	15.5 (13.1)	18.4 (13.3)	60	0.2
MLSB80	7.9 (20.2)	4.8 (22.2)	16.2 (14.2)	18.6 (12.5)	60	0.1
MLCB35	1.9 (19.7)	−2.7 (23.0)	14.9 (12.6)	18.3 (13.7)	56	0.6
MLCB80	2.4 (19.9)	−2.2 (23.3)	15.1 (12.8)	18.4 (13.9)	56	0.6
MLSB20	6.0 (21.1)	2.0 (24.3)	16.7 (13.8)	19.7 (13.7)	56	0.2
CA35	10.0 (21.2)	7.3 (23.0)	17.5 (15.4)	19.7 (13.3)	56	0.0
CA65	10.2 (19.6)	8.1 (20.1)	16.1 (14.9)	17.7 (12.1)	56	0.0
MLSB50	11.3 (19.9)	9.4 (20.5)	16.6 (15.6)	18.2 (12.9)	56	0.0
CA50	12.4 (19.9)	11.0 (20.2)	17.2 (15.8)	18.8 (12.9)	56	0.0
MLCB50	0.4 (19.2)	−4.4 (22.8)	15.0 (11.7)	18.6 (13.4)	52	0.9
CA20	7.8 (20.3)	4.70 (22.3)	16.4 (14.0)	18.8 (12.3)	52	0.1
FLM	9.8 (22.0)	6.8 (23.2)	18.0 (15.8)	20.2 (12.7)	52	0.0
FHB	16.3 (21.6)	15.1 (21.5)	20.8 (17.1)	22.8 (12.6)	48	0.0
FNB	12.5 (22.3)	10.1 (23.3)	19.2 (16.6)	21.5 (13.1)	40	0.0
I20	−33.9 (22.3)	−48.7 (37.1)	36.3 (17.9)	50.3 (34.7)	16	0.0
I80	−47.9 (21.0)	−66.0 (38.6)	49.3 (17.6)	66.9 (36.9)	4	0.0
I65	−59.8 (21.0)	−80.3 (39.3)	60.4 (19.4)	80.7 (38.5)	4	0.0
I50	−80.4 (23.0)	−106.6 (46.9)	80.4 (23.0)	106.6 (46.9)	4	0.0
I35	−52.0 (23.2)	−71.7 (42.8)	53.2 (20.1)	72.5 (41.3)	4	0.0

Table 17 Comparison of the present results with those of previously published regression equations (combined-sex test sample, n=50)

	Estimate (kg) Mean (SD)	Raw diff (kg) Mean (SD)	|PE| (SD)	20 %
Known mass	74.6		–	–
ORBA1 A&W	53.5 (10.6)	21.1 (19.2)	28.7 (15.0)	28
ORBA1 present study	75.5 (3.2)	0.9 (18.0)	19.2 (15.0)	58
HORB A&W	47.1 (11.6)	27.5 (20.9)	35.7 (17.7)	26
HORB present study	75.7 (1.1)	−1.1 (18.4)	19.9 (15.3)	56
BPOR A&W	69.7 (13.0)	4.9 (15.7)	15.6 (12.4)	70
BPOR present study	75.1 (5.6)	−0.5 (16.2)	17.1 (13.5)	64
FHB RUFF	78.1 (8.3)	−3.5 (15.6)	18.5 (12.7)	62
FHB MCHENRY	66.7 (8.7)	7.9 (15.6)	14.9 (11.4)	68
FHB GRINE	71.5 (8.8)	3.1 (15.6)	15.4 (10.5)	68
FHB RUFF 2012	67.8 (8.9)	6.8 (15.6)	14.8 (11.2)	68
FHB Present study	73.3 (7.9)	1.3 (15.7)	16.2 (10.7)	70

levels of error in our previous study (Elliott et al. 2014). The four combined-sex FHB equations were included because they are all used regularly (e.g. Trinkaus et al. 2014). We employed only the combined-sex sample to ensure a large sample size and because sex is often difficult to attribute in fragmentary skeletal remains.

Our regression equations for orbital area (ORBA) and HORB resulted in lower rates of error and placed more individuals within ±20 % of their known mass than did Aiello and Wood's (1994) equations. In contrast, our equation for BPOR had a higher percent error and estimated fewer individuals within ±20 % of known mass than did Aiello and Wood's (1994) equation (64 % compared to 70 %). Among the FHB equations, our regression equation estimated more individuals within ±20 % of known mass than did any of the published equations, including the newest one, which is designed for broad application to Holocene populations (Ruff et al. 2012). Thus, our equations generally outperformed the best equations in the literature.

Given that the majority of the equations we generated met the criteria for accuracy and generally outperformed the best equations in the literature, the study achieved its goal of improving the estimation of body mass from skeletal remains. However, several issues must be considered.

One important consideration with respect to the results is that of age. Body mass has been argued to increase with age, particularly in females and after the fifth decade, as a result of increased fat accumulation (Holloway 1980; Ruff et al. 1991, 2005). Ageing may also alter the relationship between body mass and a variable through endosteal resorption and/or increasing cross-sectional diameters through periosteal apposition (Cooper et al. 2007; Doyle et al. 2011). In addition, past populations may not have lived as long (Robson and Wood 2008) or experienced the same age-related weight gains as modern populations, making the inclusion of older individuals in reference samples potentially unnecessary for application to archaeological or fossil groups. On the other hand, body mass may decrease with advanced age (>60) due to inactivity, cachexia, and sarcopenia (Seidell and Visscher 2000; Perissinotto et al. 2002). Age is also extremely difficult to assess once skeletal maturity is reached, particularly in populations whose growth and senescent trajectories may not be the same as modern humans (Dean et al. 2014). Stature also decreases with age, as a result of disc compression, fractures, and postural changes (Cline et al. 1989). This factor may be particularly relevant for the morphometric postcranial equations and for females, as the effect may be exaggerated due to their higher susceptibility to osteoporosis (Pothiwala et al. 2006). Accordingly, the extent to which body mass predictions could vary with age is not straightforward.

We considered the effect of age on estimation accuracy for published equations in two previous papers (Elliott et al. 2014; in press) and did not find a significant influence. To explore this effect in the present research, we ran two additional sets of analyses. The first divided the test sample into three age categories: 18–39, 40–59, and 60+, while the second restricted the test sample to individuals between 18 and 60 years (n= 46).

Supplementary Table 12 gives the sample summary for the first test, and Supplementary Tables 13 and 14 provide the cranial and postcranial results for these analyses. For the cranial variables, absolute percent prediction errors were generally lower in the 18–39-year-old age group than the 40–59-year olds, but in 11 of 12 cases, the over-60 age group had the lowest error of the three groups. This suggests that age did not bias our cranial results. For the postcranial variables, the results were mixed. Joint dimensions are thought to be unresponsive to load changes throughout adult life, while diaphyseal breadths can be affected by changes in mechanical loading, environmental stress, and activity (Ruff et al. 1991; Trinkaus et al. 1994; Lieberman et al. 2001). Consequently, one might expect variables related to joint size to show higher prediction errors in older adults than in younger individuals, with diaphyseal dimensions showing the opposite effect (Ruff et al. 1991). In our sample, this means that FHB should have estimated mass more accurately in the 18–39 age group because they are likely to be closer to their "weight at maturity" than older individuals. Conversely, diaphyseal dimensions should estimate current mass in older individuals better because they have responded to external influences. However, our results did not show this consistently. For example, in keeping with expectations, FHB resulted in lower percent prediction errors in individuals younger than 40 versus older than 40. But the equation for femoral neck breadth estimated the 18–39-year olds better than the two older groups. Cortical area indices, also expected to perform better with current weight on older adults, did not consistently return lower error rates as age increased. This suggests that age was not a major factor in our postcranial results either.

Regarding the second analysis, restricting the age range to individuals younger than 60 did not substantially change the absolute percent prediction error rates for most of the cranial and postcranial variables (Supplementary Tables 15 and 16). Several variables (FHB, I80, I65, I50) showed a small drop in |PPE|, but others (FNB, CA35) showed a slight increase when age was restricted. The differences in error rates for different variables and different regions likely relate to interrelationships between age, activity, body mass, and other factors.

Overall, the results of the two tests suggest that age does not have a consistent effect on body mass estimation from skeletal remains, at least in our sample.

While our results suggest that a number of the new equations are accurate, it must be kept in mind that the equations were generated and tested under ideal conditions: both the calibration and test samples comprised individuals of known body mass and sex, and the test sample was drawn from the sample population as the calibration sample. Thus, the results represent a *best-case scenario* and must be considered to be the upper limit of accuracy for estimating body mass from hominin skeletal remains. Despite this, failure rates were surprisingly high and the body masses of many specimens were not estimated within 20 % of their actual body masses. To reiterate, in the combined-sex sample, 6 of the 12 cranial variables failed to meet the |PE| criterion and none estimated more than 64 % of the sample within 20 % of their known mass. Ten of the 24 postcranial variables failed to meet the |PE| criterion, and only 4 variables (CA35, CA65, MLSB50, and CA50) estimated 80 % or more individuals within 20 % of their known mass. This suggests that the utility of the standard, regression-based approach to estimating body mass from skeletal remains is much more limited than the field has appreciated. Even under ideal conditions, body mass estimates are not likely to be very

accurate, and any deviation in terms of incomplete or distorted elements, sex uncertainty, or proportional differences between the reference and target groups will almost certainly result in greater error. As a result, estimating body mass using any regression equation must be undertaken very cautiously and the resulting masses considered ball-park figures at best.

Our results also challenge current assumptions regarding the way different variables perform. In relation to cranial variables, orbital height (HORB) has been argued to be the best single predictor of body mass in hominoids, including humans (Aiello and Wood 1994; Churchill et al. 2012). However, in the present study, HORB was not one of the top four estimators in any of the test samples. In fact, in the combined-sex sample, it was one of the least accurate. Orbital area (ORBA) also failed to estimate mass as reliably as previous studies have suggested (Aiello and Wood 1994; Kappelman 1996; Spocter and Manger 2007). Aiello and Wood (1994) also identified biporionic breadth (BPOR) as a reliable predictor. In the present study, this variable achieved lower rates of error than other cranial variables, but still failed to estimate more than 64 % of the combined-sex sample within ±20 % of known mass. As a result, these variables probably should not be considered the most appropriate for estimating mass.

Several expectations regarding postcranial variables were also not met. For example, femoral head breadth was not among the top postcranial predictors for any of the samples and none of the FHB equations estimated more than 70 % of any of the three test samples within ±20 % of known mass. This was surprising since femoral head breadth is the most widely used skeletal variable for estimating body mass (Ruff et al. 1991; McHenry 1992; Grine et al. 1995; Ruff 2010; Ruff et al. 2012), largely because the femur bears the brunt of the body's weight. Femoral head breadth has also been argued to be more appropriate for body mass estimation than other areas of the femur because it is less responsive to external influences like environmental stresses and activity (Ruff et al. 1997). Habitual activity, in particular, is thought to influence femoral cross-sectional dimensions strongly (Ruff and Hayes 1983; Ruff and Hayes 1983; Trinkaus et al. 1991; Trinkaus and Ruff 1999). Despite this, several shaft dimensions produced lower rates of error than femoral head breadth in the present study. Consequently, these results suggest that FHB may not be as reliable an estimator of mass as it is usually assumed to be.

Although the extent to which diaphyseal dimensions are influenced by activity, environment, body mass, or some combination of these factors continues to be poorly understood (Pearson and Lieberman 2004; Pearson et al. 2008), our results suggest that shaft measurements should be investigated more thoroughly for their ability to estimate individual mass reliably. As noted, mid-shaft dimensions, particularly cortical area indices and medio-lateral shaft breadths, consistently performed better than FHB. While obtaining cortical dimensions was a difficult task in the past as it required the use of two-dimensional radiography or physical sectioning of the bone (e.g. Ruff and Hayes 1983), technologies like CT and MRI are becoming more accessible for bioanthropological research and offer the potential for accurate and non-destructive ways of accessing these data (Thali et al. 2003). Consequently, further research in this area should be pursued.

More surprising still, the results did not consistently support expectations regarding the relative reliability of cranial versus postcranial variables. Although functional relationships are not a prerequisite for good predictability (Smith 2002), most researchers argue that postcranial features will estimate mass better than cranial features because they bear the body's weight (e.g. Jungers 1990; Ruff et al. 1991). Our results suggest that this assumption needs to be examined more closely. Using the average of the 20 % criterion results for each set of variables in each sample (Table 18), the equations derived from the postcranial variables estimate mass more accurately than the cranial variables in the female and combined-sex samples. However, in males, the postcranial variables were less accurate than the cranial variables. This suggests that the relationship between lower limb morphology and mass may be different in males than in females. While this may be due to variations in activity or muscle mass, as noted earlier, the relationship between skeletal morphology, activity, and muscling continues to be a complex problem (Stirland 1998; Weiss et al. 2012; Takigawa 2014). Although accurate data regarding activity patterns and muscle mass are difficult to obtain outside specialized research settings (Kim et al. 2002), the results obtained here argue strongly in favour of further research in this area.

Table 18 Comparison of the 20 % criterion results for cranial and postcranial variables

	Females		Males		Combined sex	
	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
Cranial variables (n=12)	52.7 (2.3)	48–56	64.5 (5.0)	60–76	58.5 (2.7)	56–64
PC variables (n=19)	62.1 (17.5)	52–80	56.5 (6.2)	40–65	67.3 (10.2)	54–82
All variables	58.5 (14.4)	48–80	59.6 (6.9)	40–76	63.9 (9.2)	54–82

Lastly, our results suggest that the way in which equations are assessed for reliability may be problematic. When deriving predictive equations, most studies refer to correlation coefficients for determining which variables will be the most appropriate (Steudel 1980; Ruff et al. 1991; Delson et al. 2000; Spocter and Manger 2007, Niskanen and Junno 2009). The assumption here is that the higher the correlation coefficient, the better the variable relates to body mass and the better the resulting predictive equation will be. However, Smith (1984: 155) has argued that "a high correlation coefficient does not ensure that the corresponding regression will have good predictability". This is particularly true when sample sizes are small (Steudel 1985) and may explain the poor accuracy of the equations generated in previous studies. For example, using an interspecific sample of just five specimens, all but one of Spocter and Manger's (2007) 15 cranial variables were associated with correlation coefficients (r) greater than 0.97. With a sample size of 12, all 15 cranial variables used in Aiello and Wood's (1994) hominoid regressions had correlation coefficients greater than 0.73. On this basis, one would be led to believe that cranial variables should be good predictors of mass. However, this was not borne out by the present study or other tests (Elliott et al. 2014).

In the present study, lower correlations did not necessarily indicate poorer predictability. For example, in the combined-sex sample, the correlation coefficient between femoral head breadth (FHB) and mass was 0.40 ($p<0.01$) but 70 % of the sample was estimated within ±20 % of known mass. In contrast, femoral neck breadth correlated with mass at $r=0.43$ ($p<0.01$), yet only 56 % of the sample was estimated within ±20 % of known mass. Interestingly, these correlations are similar to those in Ruff et al. (1991), where femoral head breadth and femoral neck breadth correlated with current weight in the combined-sex sample at $r=0.49$ and $r=0.53$, respectively. Although interspecific correlation coefficients are expected to be higher than intraspecific ones (Smith 2002), these low correlation coefficients would argue against femoral head breadth as the more appropriate estimator of mass. Regardless, it is clear that correlation coefficients may not be a reliable means of assessing the predictive competence of an equation. While this point has been made before (Smith 1984) it appears to have had little impact on practice (e.g. Jung et al. 2014; Eller et al. 2014).

As an alternative to correlation coefficients, authors like Smith (2002) have argued in favour of assessing predictive performance based on the smallest standard error of the estimate (SEE). Indeed, SEEs of 0.11 and 0.09 were the basis for the suggestion that orbital height (HORB) and orbital area (ORBA) were the most reliable predictors of hominoid mass, respectively (Aiello and Wood 1994). However, SEEs are heavily influenced by sample size, with the distribution of SEE values narrowing as n increases (Hennig and Cooper 2011). Under such conditions, lower SEEs may not result in better predictability. In Aiello and Wood's (1994) case, ORBA had a lower SEE, but estimated fewer individuals within ±20 % of *known* mass than HORB (50 vs. 63 %). With our larger sample of known-mass individuals, the SEEs range was narrower (0.09–0.11 in the combined-sex sample), but the number of individuals estimated within ±20 % of known mass was considerably higher (54–82 %). Consequently, it appears that SEEs may not be a sufficient means of assessing predictive competence either.

Table 19 Comparison of correlation with SEE combined-sex calibration sample ($n=203$)

	r^2	|PE|	20 % criterion
Cranial variables	−0.729	0.710	−0.054
PC variables	−0.999	−0.001	−0.053
All variables	−0.993	−0.182	0.022

To explore these issues further, we examined the relationship between SEEs, correlation coefficients, absolute prediction errors, and the percentage of individuals who are estimated within ±20 % of their known mass. Table 19 shows that although there is a reasonably strong inverse relationship between SEEs and the correlation coefficients for a given variable, neither statistic is a good indicator of estimation competence as evaluated by the percentage difference from known mass and the "±20 % criterion". Thus, in the absence of a known-mass sample, neither the correlation coefficient of the variable nor the SEE of the regression equation is adequate to determine the accuracy of the resulting estimate. Consequently, in order to ensure a regression equation is reliable, one must start with accurately measured variables combined with associated individual body masses. The resulting equation should be tested on a known, independent sample, ideally drawn from the same population as the reference group (Giancristofaro and Salmaso 2007). Subsequent validation on known-mass samples from other populations would then provide additional confidence and broader applicability. Without these steps, equations cannot be adequately assessed for their ability to estimate mass in unknown specimens. Even under these conditions, only broad estimate ranges can be expected and any inferences drawn from them must be made with caution.

Conclusions

The results of this study support Smith's (2002: 271) contention that body mass estimation is not the "simple matter" that some researchers consider it to be. The majority of the cranial and postcranial variables tested met the criteria for acceptance as estimators of mass. In addition, most of the equations returned lower rates of error than previously published

equations for the same variables. However, the acceptance criteria used in the present study were lenient and the improvements over earlier studies were generally modest. Given the vagaries of taphonomy, uncertainties of sex attribution in fragmentary skeletal remains, and the difficulty (or worse, impossibility) of ensuring a reference sample is appropriate for the target specimen, these results suggests that body mass estimation is fraught with more uncertainty than most applications acknowledge. In addition, the attempt to derive more accurate regression equations for estimating body mass revealed other problems. Specifically, the variables currently favoured for body mass estimation may not be the most reliable. Furthermore, variables with a functional association to mass (e.g. femoral head breadth) were not consistently better predictors than those without (e.g. orbital height), and the criteria currently employed to evaluate predictive competence did not assure accuracy. While some of the variables tested here show promise as predictors of mass, further research using large documented samples needs to be undertaken to address the issues that have been identified.

Acknowledgments We are very grateful to Michael Thali, Wolf Schweitzer, and the team at the Institute for Forensic Medicine in Zurich, Switzerland for access to the CT data and their invaluable assistance during its collection. We also appreciate the input and support of the members of the Human Evolutionary Studies Program, and William Jungers. Thanks also go to two reviewers and the main editor, whose comments were very constructive. ME's work on this project was funded by a Social Sciences and Humanities Research Council Graduate Student Scholarship (#767-2009-1887 3) and Simon Fraser University. MC is funded by the Canada Research Chairs Program, the Canada Foundation for Innovation, the British Columbia Knowledge Development Fund, the Social Sciences and Humanities Research Council, and Simon Fraser University.

Grant sponsors Social Sciences and Humanities Research Council, Canada Research Chairs Program, Canada Foundation for Innovation, British Columbia Knowledge Development Fund, and Simon Fraser University Human Evolutionary Studies Program are the grant sponsors.

References

Agostini GM, Ross AH (2011) The effect of weight on the femur: a cross-sectional analysis. J Forensic Sci 56(2):339–343

Aiello LC, Wood B (1994) Cranial variables as predictors of hominine body mass. Am J Phys Anthropol 95(4):409–426

Arsuaga J-L, Lorenzo C, Carretero J-M, Gracia A, Martinez I, Garcia N, Castro J-MB, Carbonell E (1999) A complete human pelvis from the Middle Pleistocene of Spain. Nature 399(6733):255–258

Auerbach BM, Ruff CB (2004) Human body mass estimation: a comparison of "morphometric" and "mechanical" methods. Am J Phys Anthropol 125(4):331–342

Barrickman NL (2008) Evolutionary relationship between life history and brain growth in anthropoid primates. Dissertation, Duke University

Buikstra J, Ubelaker DH (1994) Standards for data collection from human skeletal remains. Arkansas Archeological Society, Fayetteville

Byard RW (2012) The complex spectrum of forensic issues arising from obesity. For Sci Med Pathol 8(4):402–413

Churchill SE, Berger LR, Hartstone-Rose A, Zondo BH (2012) Body size in African Middle Pleistocene Homo. In: Reynolds SC, Gallagher A (eds) African genesis: perspectives on hominin evolution. Cambridge University Press, Cambridge, pp 319–346

Cline MG, Meredith KE, Boyer JT, Burrows B (1989) Decline of height with age in adults in a general population sample: estimating maximum height and distinguishing birth cohort effects from actual loss of stature with aging. Hum Biol 61(3):415–425

Cohen MN, Crane-Kramer GMM (2007) Ancient health: skeletal indicators of agricultural and economic intensification. University Press of Florida, Gainesville

Cooper DM, Thomas CDL, Clement JG, Turinsky AL, Sensen CW, Hallgrímsson B (2007) Age-dependent change in the 3D structure of cortical porosity at the human femoral midshaft. Bone 40(4):957–965

Dagosto M, Terranova C (1992) Estimating the body size of Eocene primates: a comparison of results from dental and post-cranial variables. Int J Primatol 13(3):307–344

Damuth J, MacFadden B (1990) Body size in mammalian paleobiology. Cambridge University Press, Cambridge

Dean MC, Liversidge HM, Elamin F (2014) Combining radiographic and histological data for dental development to compare growth in the past and the present. Ann Hum Biol 41(4):336–347

Decker SJ, Davy-Jow SL, Ford JM, Hilbelink DR (2011) Virtual determination of sex: metric and nonmetric traits of the adult pelvis from 3D computed tomography models. J For Sci 5(56):1107–1114

Delson E, Terranova CJ, Jungers WL, Sargis EJ, Jablonski NG, Dechow PC (2000) Body mass in Cercopithecidae (Primates, Mammalia): estimation and scaling in extinct and extant taxa. American Museum of Natural History, New York

DeSilva JM (2011) A shift toward birthing relatively large infants early in human evolution. Proc Natl Acad Sci U S A 108(3):1022–1027

Doyle LE, Lazenby RA, Pfeiffer S (2011) Cortical bone mass and geometry: age, sex, and intraskeletal variation in nineteenth-century Euro-Canadians. Am J Hum Biol 23(4):534–545

Eller AR, Delson E, Guthrie EH, Frost SR (2014) Measurement protocol considerations for the cercopithecid appendicular skeleton: body mass and function. Am J Phys Anthropol 153(S58):113–113

Elliott M, Kurki H, Weston DA, Collard M (2014) Estimating fossil hominin body mass from cranial variables: an assessment using CT data from modern humans of known body mass. Am J Phys Anthropol 154(2):201–214

Elliott M, Kurki H, Weston DA, Collard M (2015) Estimating body mass from postcranial variables: an evaluation of current equations using a large known-mass sample of modern humans. Archaeol Anthropol Sci. doi:10.1007/s12520-015-0251-6

Frayer D (1984) Biological and cultural change in the European Late Pleistocene and Early Holocene. In: Smith F, Spencer F (eds) The origins of modern humans: a world survey of the fossil evidence. Wiley-Liss, New York, pp 211–250

Giancristofaro RA, Salmaso L (2007) Model performance analysis and model validation in logistic regression. Statistica 63(2):375–396

Grine FE, Jungers WL, Tobias PV, Pearson OM (1995) Fossil *Homo* femur from Berg Aukas, northern Namibia. Am J Phys Anthropol 97(2):151–185

Harrell FE, Lee KL, Mark DB (1996) Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15(4):361–387

Hennig C, Cooper D (2011) Brief communication: the relation between standard error of the estimate and sample size of histomorphometric aging methods. Am J Phys Anthropol 145(4):658–664

Hinton P (2004) Statistics explained, 2nd edn. Routledge, New York

Holloway RL (1980) Within-species brain-body weight variability: a re-examination of the Danish data and other primate species. Am J Phys Anthropol 53(1):109–121

Jung GU, Lee UY, Kim DH, Kwak DS, Ahn YW, Kim YS (2014) The body mass estimation from human talus: inductive approach and 3D morphometric study. FASEB J 28(S1):919.1

Jungers WL (1990) Scaling of hominoid femoral head size and the evolution of hominid bipedalism. Am J Phys Anthropol 33(S11):246

Kappelman J (1996) The evolution of body mass and relative brain size in fossil hominids. J Hum Evol 30(3):243–276

Kim J, Wang Z, Heymsfield SB, Baumgartner RN, Gallagher D (2002) Total-body skeletal muscle mass: estimation by a new dual-energy X-ray absorptiometry method. Am J Clin Nutr 76(2):378–383

Kim G, Jung HJ, Lee HJ, Lee JS, Koo S, Chang SH (2012) Accuracy and reliability of length measurements on three-dimensional computed tomography using open-source OsiriX software. J Digit Imaging 25(4):486–491

Konigsberg LW, Hens SM, Jantz LM, Jungers WL (1998) Stature estimation and calibration: Bayesian and maximum likelihood perspectives in physical anthropology. Am J Phys Anthropol 107(S27):65–92

Kordos L, Begun DL (2001) Primates from Rudabanya: allocation of specimens to individuals, sex and age categories. J Hum Evol 40(1):17–39

Kurki HK, Ginter JK, Stock JT, Pfeiffer S (2010) Body size estimation of small-bodied humans: applicability of current methods. Am J Phys Anthropol 141(2):169–180

Lieberman DE, Devlin MJ, Pearson OM (2001) Articular area responses to mechanical loading: effects of exercise, age, and skeletal location. Am J Phys Anthropol 116(4):266–277

Lieberman DE, Polk JD, Demes B (2004) Predicting long bone loading from cross-sectional geometry. Am J Phys Anthropol 123(2):156–171

Lorkiewicz-Muszyńska D, Przystańska A, Kociemba W, Sroka A, Rewekant A, Żaba C, Paprzycki W (2013) Body mass estimation in modern population using anthropometric measurements from computed tomography. Forensic Sci Int 231(1–3):405-e1–6

McHenry HM (1992) Body size and proportions in early hominids. Am J Phys Anthropol 87(4):407–431

Myszka A, Piontek J, Vančata A (2012) Body mass reconstruction on the basis of selected skeletal traits. Anthropol Anz 69(3):305–315

Niskanen M, Junno JA (2009) Estimation of African apes' body size from postcranial dimensions. Primates 50(3):211–220

Pearson OM, Cordero RM, Busby AM (2008) How different were Neanderthals' habitual activities? A comparative analysis with diverse groups of recent humans. In: Hublin JJ, Harvati K, Harrison T (eds) Neanderthals revisited: new approaches and perspectives. Springer, Netherlands, pp 135–156

Pearson OM, Lieberman DE (2004) The aging of Wolff's "law": ontogeny and responses to mechanical loading in cortical bone. Am J Phys Anthropol 125(S39):63–99

Perissinotto E, Pisent C, Sergi G, Grigoletto F, Enzi G, Group IW (2002) Anthropometric measurements in the elderly: age and gender differences. Br J Nutr 87(2):177–186

Pomeroy E, Stock JT (2012) Estimation of stature and body mass from the skeleton among coastal and mid-altitude Andean populations. Am J Phys Anthropol 147(2):264–279

Porter AMW (1999) The prediction of physique from the skeleton. Int J Osteoarchaeol 9(2):102–115

Pothiwala P, Evans EM, Chapman-Novakofski KM (2006) Ethnic variation in risk for osteoporosis among women: a review of biological and behavioral factors. J Women Health 15(6):709–719

R Core Team (2010) R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org

Rainwater C, Cabo-Perez L, Symes S (2007) Body mass estimation and personal identification. Am J Phys Anthropol 134(S45):194–195

Rightmire GP (2004) Brain size and encephalization in early to mid-Pleistocene Homo. Am J Phys Anthropol 124(2):109–123

Robson SL, Wood B (2008) Hominin life history: reconstruction and evolution. J Anat 212(4):394–425

Ruff CB (1991) Climate and body shape in hominid evolution. J Hum Evol 21(2):81–105

Ruff CB (1994) Morphological adaptation to climate in modern and fossil hominids. Am J Phys Anthropol 37(S19):65–107

Ruff CB (2000a) Body mass prediction from skeletal frame size in elite athletes. Am J Phys Anthropol 113(4):507–517

Ruff CB (2000b) Body size, body shape, and long bone strength in modern humans. J Hum Evol 38(2):269–290

Ruff CB (2002) Variation in human body size and shape. Ann Rev Anthropol 31:211–232

Ruff CB (2010) Body size and body shape in early hominins—implications of the Gona pelvis. J Hum Evol 58(2):166–178

Ruff CB, Hayes W (1983) Cross-sectional geometry of Pecos Pueblo femora and tibiae—a biomechanical investigation: I. Method and general patterns of variation. Am J Phys Anthropol 60(3):359–381

Ruff CB, Holt BM, Niskanen M, Sladék V, Berner M, Garofalo E, Garvin HM, Hora M, Maijanen H, Niinimäki S et al (2012) Stature and body mass estimation from skeletal remains in the European Holocene. Am J Phys Anthropol 148(4):601–617

Ruff CB, Niskanen M, Junno J-A, Jamison P (2005) Body mass prediction from stature and bi-iliac breadth in two high latitude populations, with application to earlier higher latitude humans. J Hum Evol 48(4):381–392

Ruff CB, Scott WW, Liu AYC (1991) Articular and diaphyseal remodeling of the proximal femur with changes in body mass in adults. Am J Phys Anthropol 86(3):397–413

Ruff CB, Trinkaus E, Holliday TW (1997) Body mass and encephalization in Pleistocene *Homo*. Nature 387(6629):173–176

Ruff CB, Walker A (1993) Body size and body shape. In: Walker A, Leakey R (eds) The Nariokotome *Homo erectus* skeleton. Harvard University Press, Cambridge, pp 234–265

Seidell JC, Visscher TL (2000) Body weight and weight change and their health implications for the elderly. Eur J Clin Nutr 54(S3):S33–S39

Smith RJ (1984) Allometric scaling in comparative biology: problems of concept and method. Am J Physiol Regul Integr Comp Physiol 246(2): R152–R160

Smith RJ (1993) Logarithmic transformation bias in allometry. Am J Phys Anthropol 90(2):215–228

Smith RJ (2002) Estimation of body mass in paleontology. J Hum Evol 43(2):271–287

Smyth AM, Viner MD, Conlogue GJ, Brownlee SA, Aronsen GP (2012) An evaluation of medical imaging techniques for craniometric data collection. Am J Phys Anthropol 147(S54):274–274

Spocter MA, Manger PR (2007) The use of cranial variables for the estimation of body mass in fossil hominins. Am J Phys Anthropol 134(1):92–105

Steckel RH, Rose JCT (2002) The backbone of history: health and nutrition in the western hemisphere. Cambridge University Press, New York

Stirland AJ (1998) Musculoskeletal evidence for activity: problems of evaluation. Int J Osteoarchaeol 8(5):354–362

Steudel K (1980) New estimates of early hominid body size. Am J Phys Anthropol 52(1):63–70

Steudel K (1985) Allometric perspectives on fossil catarrhine morphology. In: Jungers WJ (ed) Size and scaling in primate biology. Springer, New York, pp 449–475

Suskewicz JA (2004) Estimation of living body weight based on measurements of anterior superior iliac spine breadth and stature. Masters Thesis, Louisiana State University

Swiss Federal Statistical Office (2012) Population size and composition. http://www.bfs.admin.ch/bfs/portal/en/index.html

Takigawa W (2014) Age changes of musculoskeletal stress markers and their inter-period comparisons. Anthropol Sci 122(1):7–22

Thali MJ, Yen K, Schweitzer W, Vock P, Boesch C, Ozdoba C, Schroth G, Ith M, Sonnenschein M, Doernhoefer T et al (2003) Virtopsy, a new imaging horizon in forensic pathology: virtual autopsy by postmortem multislice computed tomography (MSCT) and magnetic resonance imaging (MRI): a feasibility study. J Forensic Sci 48(2):386–403

Trinkaus E, Jelínek J (1997) Human remains from the Moravian Gravettian: the Dolní Věstonice 3 postcrania. J Hum Evol 33(1):33–82

Trinkaus E, Ruff CB (1999) Diaphyseal cross-sectional geometry of Near Eastern Middle Paleolithic humans: the femur. J Archaeol Sci 26(4): 409–424

Trinkaus E, Churchill SE, Ruff CB (1994) Postcranial robusticity in *Homo*. II: Humeral bilateral asymmetry and bone plasticity. Am J Phys Anthropol 93(1):1–34

Trinkaus E, Churchill SE, Villemeur I, Riley KG, Heller JA, Ruff CB (1991) Robusticity versus shape: the functional interpretation of Neandertal appendicular morphology. Zinruigaku zassi 99(3):257–278

Trinkaus E, Buzhilova AP, Mednikova MB, Dobrovolskaya MV (2014) The people of Sunghir: burials, bodies, and behavior in the earlier Upper Paleolithic. Oxford University Press, New York

Weiss E, Corona L, Schultz B (2012) Sex differences in musculoskeletal stress markers: problems with activity pattern reconstructions. Int J Osteoarchaeol 22(1):70–80

Wood BA, Collard M (1996) Comments and reply to Smith: biology and body size in human evolution: statistical inference misapplied. Curr Anthropol 37(3):451–481

Wood BA, Collard M (1999) The human genus. Science 284(5411): 65–71

Wu G, Baraldo M, Furlanut M (1995) Calculating percentage prediction error: a user's note. Pharmacol Res 32(4):241–248

Wu X, Athreya S (2013) Description of the geological context, discrete traits, and linear morphometrics of the Middle Pleistocene hominin from Dali, Shaanxi Province, China. Am J Phys Anth. 150(1):141–157