AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 154:201–214 (2014)

Estimating Fossil Hominin Body Mass From Cranial Variables: An Assessment Using CT Data From Modern Humans of Known Body Mass

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

[1]*Human Evolutionary Studies Program and Department of Archaeology, Simon Fraser University, Burnaby, BC, Canada V5A 1S6*
[2]*Department of Anthropology, University of Victoria, Victoria, BC, Canada V8W 2Y2*
[3]*Department of Anthropology, University of British Columbia, Vancouver, BC, Canada V6T 1Z1*
[4]*Department of Human Evolution, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany*
[5]*Department of Archaeology, University of Aberdeen, Aberdeen, Scotland AB24 3UF, UK*

KEY WORDS body mass; hominin; paleoanthropology; osteology

ABSTRACT Body mass estimates are integral to a wide range of inferences in paleoanthropology. Most techniques employ postcranial elements, but predictive equations based on cranial variables have also been developed. Three studies currently provide regression equations for estimating mass from cranial variables, but none of the equations has been tested on samples of known mass. Nor have the equations been compared to each other in terms of performance. Consequently, this study assessed the performance of existing cranial equations using computed tomography scans from a large, documented sample of modern humans of known body mass. Virtual models of the skull were reconstructed and measured using computer software, and the resulting variables were entered into three sets of published regression equations. Estimated and known body masses were then compared. For most equations, prediction errors were high and few individuals were estimated within ±20% of their known mass. Only one equation satisfied the accuracy criteria. In addition, variables that had been previously argued to be good predictors of mass in hominins, including humans, did not estimate mass reliably. These results have important implications for paleoanthropology. In particular, they emphasize the need to develop new equations for estimating fossil hominin body mass from cranial variables. Am J Phys Anthropol 154:201–214, 2014. © 2014 Wiley Periodicals, Inc.

The estimation of body mass from skeletal remains continues to be a crucial task for paleoanthropologists. Body mass has been shown to affect a wide range of ecological, behavioral and life history traits (Calder, 1984; Damuth and MacFadden, 1990; Smith, 1996). Consequently, estimates of body mass are used regularly to infer the characteristics of fossil hominins (e.g., Wood and Collard, 1999; Ruff, 2002; Sciulli and Blatt, 2008; Kurki et al., 2010; Churchill et al., 2012). They are also used to control for the confounding effects of body size differences in comparative analyses, particularly of features like brain size, tooth size, and limb proportion (e.g., Smith and Jungers, 1997; Rightmire, 2004; DeSilva and Lesnik, 2008).

Numerous equations for estimating fossil hominin body mass have been developed over the last 25 years (Ruff et al., 1991, 1997, 2012; McHenry, 1992; Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007). Obtained by regressing a skeletal variable on body mass in samples of extant taxa, these equations generally have the form $Y = a + bX$, where Y is the estimated mass, X is the skeletal measurement, and a and b are the intercept and slope of the regression line, respectively. While most employ postcranial measurements (Ruff et al., 1991, 1997, 2012; McHenry, 1992), equations based on cranial measurements have also been developed (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007). The postcranial equations use either femoral head breadth (Ruff et al., 1991, 1997, 2012; McHenry, 1992; Grine et al., 1995) or a combination of stature and bi-iliac breadth (Ruff, 1991; Ruff et al., 1997, 2005) and are based on values for individual modern humans from a single population (e.g., Ruff et al., 1991), or on mean

---

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Social Sciences and Humanities Research Council Graduate Student Scholarship (to ME); Grant number: #767-2009-1887 3; Grant sponsors: Simon Fraser University (to ME), Canada Research Chairs Program, Canada Foundation for Innovation, British Columbia Knowledge Development Fund, the Social Sciences and Humanities Research Council, and Simon Fraser University (to MC).

*Correspondence to: Mark Collard, Human Evolutionary Studies Program and Department of Archaeology, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6.
E-mail: mcollard@sfu.ca

Received 18 June 2013; revised 29 January 2014; accepted 29 January 2014

DOI: 10.1002/ajpa.22493

data for multiple modern human groups (e.g., Ruff, 1991; Ruff et al., 2005). The cranial equations employ a range of cranial variables, including orbital height, orbital area, and biporionic breadth (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007) and are derived from means for multiple extant primate species, including modern humans.

The accuracy of the postcranial equations has been assessed several times (Ruff et al., 1997, 2012; Ruff, 2000; Auerbach and Ruff, 2004; Lorkiewicz-Muszyńska et al., 2013). However, the cranial equations have not been tested with independent data. They have been argued to be valid because they produce estimates that broadly agree with the estimates yielded by the postcranial equations (Aiello and Wood, 1994), but their accuracy has never been formally evaluated. In addition, the various sets of cranial equations have never been compared to each other, as has been done with the postcranial equations (Auerbach and Ruff, 2004). Given that body mass estimates obtained with the cranial equations are used to inform theories concerning human evolution on a regular basis (e.g., Wood and Collard, 1999; McHenry and Coffing, 2000; Aiello and Key, 2002; Churchill et al., 2012), there was a clear need to assess their reliability.

In view of the foregoing, the present study used computerized tomography (CT)-derived cranial data from a large sample of modern humans of known body mass to assess the accuracy of published cranial equations (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007). Each of these publications provides equations derived from both a broad "all primate" sample and a narrower sample consisting only of apes and humans. For estimating fossil hominin body mass, however, all three recommend using the equations derived from the narrower, hominoid sample (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007). Accordingly, this study focused on the hominoid-only equations.

The primary goals of this study was to test the way in which cranial measurements are used to estimate fossil hominin body mass in paleoanthropology. As the equations in question are derived from data sets consisting of means for multiple extant primate species, it is sometimes assumed that they are used to estimate the body mass of species, rather than single individuals within a species (Hartwig-Scherer and Martin, 1992). However, this is not the case. The studies that derived the equations used them specifically to estimate the body masses of individual fossil hominin specimens (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007). The equations have also been applied to other fossil primate individuals (e.g., Kordos and Begun, 2001). Accordingly, we used the cranial equations to estimate the body masses for each individual in our sample, as would be done with a single fossil specimen. These estimates were then compared to the individuals' known body masses.

In general, regression-based equations for predicting body mass are expected to work best when applied to specimens whose taxon is included in the sample used to generate the equations (Smith, 2002). As all the samples used to generate the cranial equations contained modern humans (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007), the use of a modern human sample in the present study was expected to provide a reasonable baseline for the likely accuracy of the equations as applied to fossil hominin specimens.

MATERIALS AND METHODS

Sample

This study used archived CT scan data from a sample of 253 deceased modern human adults. The sample consisted of 128 males and 125 females, between 18 and 90 years (M mean = 48.1 years, F mean = 51.2 years). The data were obtained from the Institute of Forensic Medicine (IFM) at the University of Zurich, Switzerland where whole-body CT scans are routinely taken for all individuals entering the facility for forensic evaluation (Thali et al., 2007). The scans are maintained on the IFM's secure server, and were accessed with the approval of the IFM in accordance with its protocols.

Sample individuals were selected through query searches of the IFM's database, record review, and visual inspection of the CT scans. Individuals with skeletal abnormalities, trauma, or cranial implants were excluded, as were individuals who were processed more than three days after death. Sex, age at death (in years), body mass (in kg), and stature (in cm) were recorded for each individual. Body mass index (BMI) was calculated from body mass and stature using the standard equation (mass/stature^2) to provide an indication of overall body condition. As population affinity is not recorded on postmortem documentation in Switzerland it was not included as a variable in the present study. However, as more than 80% of the Swiss population is of European descent (SFSO, 2012), the sample was considered European. Table 1 provides the summary data for the sample.

Imaging and three-dimensional reconstruction

Deceased individuals entering the IFM for forensic evaluation are scanned using a 128-slice, Siemens SOMATOM® Definition Flash, Dual-source CT scanner (Siemens Healthcare, Forchheim, Germany). Scans of the whole body, as well as specific areas of interest (e.g., head and chest), are taken at 120 kV with mAs and field of view adjusted for optimal resolution. Cranial data are reconstructed with slice thicknesses of 0.75 mm (0.375

TABLE 1. Summary data for test sample

	Females (n = 125)			Males (n = 128)			Combined sample (n = 253)		
Variable	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Weight (kg)	69.5	19.3	31.8–146.0	81.6	16.4	40.5–142.2	75.6	18.8	31.8–146.0
Stature (cm)	166.3	8.2	149.0–195.0	177.5	7.9	154.0–193.0	171.9	9.8	149.0–195.0
Age (years)	51.2	16.5	18.0–90.0	48.1	14.1	18.0–80.0	49.6	15.3	18.0–90.0
BMI[a]	25.1	6.4	14.3–46.5	25.8	4.6	15.4–46.9	25.4	5.6	14.3–46.9

[a] BMI, body mass index, calculated as mass (kg)/[stature (m)]^2.

mm overlap), using bone convolution kernels. All data are archived as Digital Information and Communications in Medicine (DICOM) files on the IFM's secure Picture Archiving and Communication System (PACS) server (IDS7, Version 12.2.3.297, 2010, Sweden).

CT scan data for each patient were accessed specifically for this project from the IFM PACS server using OsiriX imaging software (http://www.osirix-viewer.com). Scans were anonymized and three-dimensional (3D) virtual models were volume rendered from the DICOM slice data using presets provided by OsiriX. Crania were then oriented in consistent planes (coronal, sagittal, or transverse) for visualization and measurement. Measurements were taken on the right side, to the nearest 0.1 mm using OsiriX tools. The accuracy of 3D volume rendered models from CT has been demonstrated previously in a number of studies (Cavalcanti et al., 2004; Lopes et al., 2008; Decker et al., 2011; Smyth et al., 2012). To verify this for the present study, an archaeological skull from the IFM's collection was measured using standard calipers. It was then scanned, virtually reconstructed, and remeasured according to the method outlined above. Differences between the measurements recorded on the physical and virtual skulls were less than 3% for all variables.

Variables

The variables selected for this study were chosen on the basis of their performance in previous analyses. Specifically, Aiello and Wood (1994) identified orbital area, orbital height, and biporionic breadth as good predictors of body mass in their hominoid-only sample. Kappelman (1996) also found orbital area and orbital height to be strongly correlated with body mass in hominoids. Spocter and Manger (2007) identified foramen magnum area, foramen magnum area calculated as an ellipse, and biorbital breadth as the best predictors of body mass in their hominoid sample. They also found orbital height, orbital area, orbital area as an ellipse, and biporionic breadth to be good predictors of body mass in hominoids.

Thus, a total of six linear measurements were taken for this study: orbital height, orbital breadth, biorbital breadth, foramen magnum length, foramen magnum breadth, and biporionic breadth (Table 2; Fig. 1). Intraobserver repeatability for these measurements was tested by remeasuring them on nine randomly selected crania, with a 3-week time lapse. Mean percentage errors were all below 0.5%.

Because orbital area was not calculated in the same way by Aiello and Wood (1994), Kappelman (1996), and Spocter and Manger (2007), three orbital area calculations were also included in the present study. The first orbital area (ORBA1) involved a simple breadth–height calculation (area $= b \times h$) following the method used in Aiello and Wood (1994) and Spocter and Manger (2007). The second orbital area (ORBA2), was calculated as an ellipse [area $= (\pi/4) \times b \times h$]. This followed Spocter and Manger (2007) who argued that it provided a more accurate reflection of true area than simple breadth × height. The third orbital area (ORBA3) employed a method similar to Kappelman (1996) in which 2D images of the orbit were imported into a computer-aided design (CAD) program (in this case, ImageJ v.1.46, rsbweb.nih.gov/ij/) and the area measured by tracing the margin perimeter and using the "area" function of the program.

Finally, two foramen magnum areas were included for the same reason. The first area, (FMA1) was calculated as breadth × length in the same way as Aiello and Wood (1994) and Spocter and Manger (2007). The second area (FMA2) was calculated as an ellipse following Spocter and Manger's (2007) study. Summary data for the linear measurements and area calculations are provided in the Supporting Information Table S1.

TABLE 2. Cranial variables

No.	Abbreviation	Description	References[a]
1	BORB	Breadth of orbit: distance between maxillofrontale and ectoconchion in mm	AW[1], SM[10], M[51]
2	HORB	Height of orbit: distance between superior and inferior orbital margins, taken at a right angle to BORB in mm	AW[2], K[1], SM[9], M[52]
3	BIOR	Biorbital breadth: distance between two ectoconchion in mm	AW[5], SM[8], M[44]
4	BPOR	Biporionic breadth: distance from porion to porion in mm	AW[7], SM [13]
5	LFM	Length of foramen magnum: distance between basion and opisthion in mm	AW[10], SM[1], M[8]
6	BFM	Breadth of foramen magnum: distance in the coronal plane between the inner margins of the foramen magnum in mm	AW[11], SM[2], M[16]
7	ORBA1	Orbital area ($b \times h$): product of breadth × height in mm^2	AW[3], SM[11]
8	ORBA2	Orbital area (ellipse): calculated from breadth × height as an ellipse in mm^2	SM[12]
9	ORBA3	Orbital area (CAD): calculated from perimeter margin using area function of ImageJ in mm^2	K[2]
10	FMA1	Foramen magnum area ($b \times l$): product of breadth × length in mm^2	AW[12], SM[3]
11	FMA2	Foramen magnum area (ellipse): calculated from breadth × length as an ellipse in mm^2	SM[4]

[a] Literature sources: AW, Aiello and Wood (1994); K, Kappelman (1996); SM, Spocter and Manger (2007), square brackets refer to original variable number. As several measurements also correspond to those in Martin (1928), the variable references from that publication are also included (e.g., M[44]).

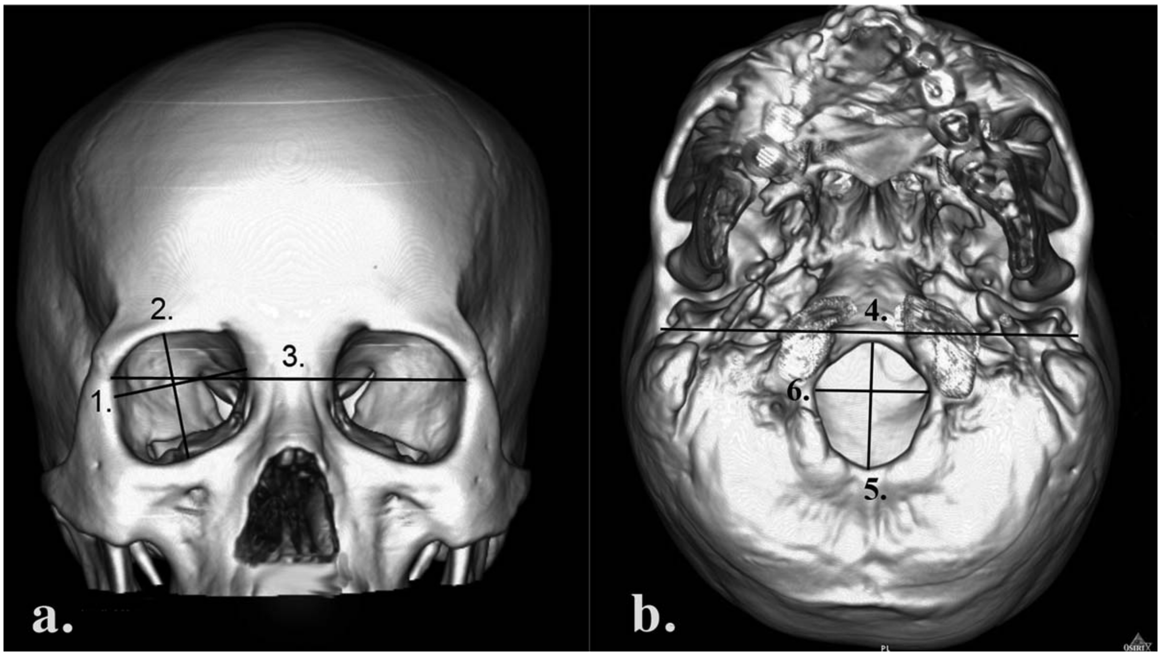


Fig. 1. Frontal (**a**) and basicranial (**b**) example of skull three-dimensionally rendered from CT data, showing linear variables used for this study.

TABLE 3. Published LSR regression equations for body mass estimation (hominoids)

Variable	Aiello and Wood (1994)	CF	Kappelman (1996)	CF	Spocter and Manger (2007)	CF
BORB	$5.22 \times \text{BORB} - 3.35$	1.055	n/a	n/a	$3.78 \times \text{BORB} - 1.31$	1.055
HORB	$4.42 \times \text{HORB} - 2.12$	1.025	$4.45 \times \text{HORB} - 2.16$	1.048	$4.45 \times \text{HORB} - 2.64$	1.0
BIOR	$4.82 \times \text{BIOR} - 4.67$	1.045	n/a	n/a	$3.81 \times \text{BIOR} - 3.29$	1.025
BPOR	$3.77 \times \text{BPOR} - 2.95$	1.04	n/a	n/a	$4.82 \times \text{BPOR} - 4.92$	0.99
LFM	$3.07 \times \text{LFM} + 0.18$	1.1	n/a	n/a	$3.86 \times \text{LFM} - 1.24$	1.03
BFM	$3.74 \times \text{BFM} - 0.48$	1.09	n/a	n/a	$3.77 \times \text{BFM} - 0.73$	1.025
ORBA1	$2.47 \times \text{ORBA1} - 2.92$	1.025	n/a	n/a	$2.16 \times \text{ORBA1} - 2.27$	1.025
ORBA2	n/a	n/a	n/a	n/a	$4.34 \times \text{ORBA2} - 5.79$	1.01
ORBA3	n/a	n/a	$2.26 \times \text{ORBA3} - 2.18$	1.025	n/a	n/a
FMA1	$1.70 \times \text{FMA1} - 0.16$	1.09	n/a	n/a	$1.93 \times \text{FMA1} - 1.03$	1.03
FMA2	n/a	n/a	n/a	n/a	$3.82 \times \text{FMA2} - 4.06$	1.02

All data log (base10) transformed. Correction factor (CF) is mean of Smearing and Ratio estimates taken from each study. Measurements in mm or mm^2, resulting BM in g.

TABLE 4. Published RMA regression equations for body mass estimation (hominoid)

Variable	Aiello and Wood (1994)	Spocter and Manger (2007)
BORB	$5.46 \times \text{BORB} - 3.7$	$4.09 \times \text{BORB} - 1.77$
HORB	$4.53 \times \text{HORB} - 2.29$	$4.48 \times \text{HORB} - 2.69$
BIOR	$5.1 \times \text{BIOR} - 5.2$	$3.88 \times \text{BIOR} - 3.43$
BPOR	$3.84 \times \text{BPOR} - 3.1$	$4.91 \times \text{BPOR} - 5.09$
LFM	$3.4 \times \text{LFM} - 0.28$	$3.94 \times \text{LFM} - 1.37$
BFM	$4.06 \times \text{BFM} - 0.89$	$3.88 \times \text{BFM} - 0.87$
ORBA1	$2.52 \times \text{ORBA1} - 3.05$	$2.19 \times \text{ORBA1} - 2.39$
ORBA2	n/a	$4.39 \times \text{ORBA2} - 5.92$
FMA1	$1.86 \times \text{FMA1} - 0.58$	$1.96 \times \text{FMA1} - 1.14$
FMA2	n/a	$3.88 \times \text{FMA} - 4.19$

All data log (base10) transformed. Measurements in mm or mm^2, resulting BM in g.

Analyses

As noted previously, this study focused on the putatively most accurate of the equations derived from the hominoid-only samples. Within these, Aiello and Wood (1994) and Spocter and Manger (2007) provide equations for both least square regression (LSR) and reduced major axis (RMA) techniques. Kappelman (1996) provides only LSR-based equations for the two variables used in his study. Table 3 lists the LSR-based equations tested, while Table 4 lists the RMA-based equations. For each set of equations, analyses were carried out on the full sample of 253 individuals, as well as subsamples of females ($n = 125$) and males ($n = 128$).

The choice of line-fitting technique is an important consideration when generating predictive equations via regression, and much has been written on the relative merits of different approaches (Hartwig-Scherer and Martin, 1992; Smith, 1996, 2009; Konigsberg et al., 1998). Some researchers argue that LSR is the best method for predicting one variable from another, especially when the goal is to minimize the error of the dependent variable (Sokal and Rohlf, 1995; Smith, 1996, 2009; Konigsberg et al., 1998). Others, however, contend that LSR is problematic because it unrealistically assumes that the independent variable is sampled without error, and produces biased results when applied outside the range of the sample from which it was derived (Konigsberg et al., 1998; Ruff et al., 2012). Accordingly, they recommend RMA regression because it assumes that both variables are sampled with error and produces better results when extrapolating beyond the range of the original data set (Auerbach and Ruff, 2004). Because consensus regarding which method is better to use when predicting body mass remains elusive (Hartwig-Scherer and Martin, 1992; Smith, 2009), we tested both LSR-based and RMA-based equations when they were provided.

To assess the accuracy of the equations, linear measurements and areas were first log (Base 10) transformed and the resulting values entered into the appropriate equations. Estimated weights were then de-transformed and converted to kilograms. For the LSR-based analyses, masses were multiplied by correction factors provided by each method to account for the de-transformation process (Smith, 1996). Subsequently, raw and percentage differences between the known and estimated body masses (EBMs) were calculated for each individual. Raw differences were calculated as (known − EBM), while percentage difference was calculated using the equation for percentage of prediction error (PPE): PPE = (known − EBM)/known × 100 (Wu et al., 1995). PPEs provide the directional difference between the known and estimated masses. A positive PPE value indicates that the known mass is larger than the estimated mass and the equation underestimates mass. A negative PPE value indicates that the known mass is smaller than the estimated mass and the equation overestimates mass. PPEs were calculated for males and females as well as the combined-sex sample. Absolute percentage differences (|PPE|) were also calculated for each group to assess the magnitude of the difference between the estimated and known masses (Dagosto and Terranova, 1992; Aiello and Wood, 1994). Medians, extremes, and quartiles of the differences between known and estimated mass were also plotted to evaluate their variability and bias (cf. Pomeroy and Stock, 2012), and paired t-tests were carried out to establish the significance of these differences. Also in keeping with previous studies (Dagosto and Terranova, 1992; Aiello and Wood, 1994; Spocter and Manger, 2007), the percentage of individuals whose EBM fell within ±20% of their known mass was also calculated. As a final comparison, we calculated the raw means (in kgs) and 95% confidence intervals (CIs) for the predicted masses for each equation. However, for consistency with the existing studies, the PPEs and the "percent-within-20%" values were used as the primary criteria of assessment for each equation. Analyses were conducted in "R" (R Development Core Team, 2010).

Expectations

The validity of a predictive equation depends largely on its ability to estimate a known quantity with reasonable accuracy. However, in the case of body mass estimation, there is little consensus regarding the definition of "reasonable". For example, in considering body mass estimates for Eocene primates, Dagosto and Terranova (1992) considered mean percentage differences between known and EBM of 15–30% to be largely inaccurate. In contrast, Aiello and Wood (1994) considered several cranial variables to be reliable predictors of body mass despite prediction errors of 15–19%. Spocter and Manger (2007) also accepted variables with prediction errors of 10–16% as reliable in some of their analyses. In light of this variability, we chose to err on the side of leniency and accepted absolute prediction errors of 19% or less as our primary criterion of accuracy.

As a second criterion, we also calculated the number of individuals whose body mass fell within ±20% of their known mass. Ruff et al. (2005) have suggested that a reliable equation for estimating body mass should estimate the majority of test individuals within 10 or 15% of their known mass. However, Barrickman (2008) has argued that an equation only needs to estimate between 60% and 70% of the specimens within 20% of their known mass to be considered reliable. Other studies 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). Again, for the current study, we adopted a conservative approach and used "50% of the specimens estimated within ±20% of known mass" as the lower limit for an equation to be acceptable.

In addition to assessing the equations by the two criteria outline above, we made specific predictions regarding their performance based on the results of the original studies. Because Aiello and Wood (1994:421) considered orbital area, orbital height, and biporionic breadth to "give the most reliable predictions of body masses for hominoids, including humans", these variables were expected to perform well. In particular, orbital height was expected to perform best because its predicted mass corresponded closely with those from postcranial variables and Aiello and Wood (1994:424) recommend it as the overall "preferred cranial predictor" for large-bodied hominines.

Of the two variables he tested, Kappelman (1996) found the CAD-derived orbital area to be a better predictor of mass than orbital height. Consequently, we expected the equation for orbital area to perform better than that for orbital height.

Spocter and Manger (2007) found foramen magnum area, foramen magnum area as an ellipse, biorbital breadth and biporionic breadth to be the best predictors of mass in their hominoid-only sample. However, orbital height, orbital area and orbital area as an ellipse performed almost as well in their study and also returned low prediction errors (<13%). Consequently, all seven variables were expected to perform well in the current sample.

RESULTS

The results of the tests of the LSR-based equations are summarized in Tables (5–7) and Figures 2–4. Table 8 compares the mean predicted masses, the differences from known mean, and provides 95% confidence intervals for the predicted masses for each equation, by study source. Aiello and Wood's (1994) and Spocter and Manger's (2007) RMA-based equations consistently returned higher mean errors, and estimated fewer individuals within ±20% of their known mass than the LSR-based equations. Consequently, the results of the RMA analyses are not reported here, but are summarized in the Supporting Information Tables S2 and S3.

Aiello and Wood's (1994) equations

Aiello and Wood's (1994) equations did not estimate mass reliably according to the |PPE| and 50%-within-

TABLE 5. Difference between known and estimated mass, Aiello and Wood (1994) LSR-based equations

	Females ($n = 125$)			Males ($n = 128$)			Combined ($n = 253$)		
Variable	PPE, mean^{a} (SD)	|PPE|, mean^{b} (SD)	20% (%)	PPE, mean^{a} (SD)	|PPE|, mean^{b} (SD)	20% (%)	PPE, mean^{a} (SD)	|PPE|, mean^{b} (SD)	20% (%)
BORB	11.42 (32.1)^{c}	27.26 (20.3)	44.00	4.63 (30.2)^{c}	25.12 (17.2)	42.19	7.99 (31.3)^{c}	26.18 (18.8)	43.08
HORB	26.63 (28.9)^{c}	34.37 (19.0)	28.00	36.84 (22.8)^{c}	39.46 (17.8)	18.75	31.79 (26.4)^{c}	36.94 (18.5)	23.32
BIOR	−10.40 (34.0)	24.97 (25.2)	56.80	−15.32 (30.5)^{c}	24.24 (23.9)	53.91	−12.89 (32.3)^{c}	24.60 (24.5)	55.34
BPOR	3.33 (28.2)^{c}	21.42 (18.6)	57.60	**−1.40 (21.8)**	**17.46 (13.0)**	**64.84**	0.94 (25.2)^{c}	19.42 (16.1)	61.26
LFM	−23.59 (39.6)^{c}	33.51 (31.6)	41.60	−18.49 (34.3)^{c}	28.14 (26.9)	52.34	−21.01 (37.1)^{c}	30.80 (29.4)	47.04
BFM	−55.59 (60.9)^{c}	62.64 (53.6)	25.60	−52.37 (54.1)^{c}	56.72 (49.4)	21.86	−53.96 (57.5)^{c}	59.65 (51.5)	23.72
ORBA1	22.17 (25.8)^{c}	28.86 (18.0)	35.20	25.64 (20.8)^{c}	28.73 (16.3)	29.69	23.93 (23.4)^{c}	28.79 (17.1)	32.41
FMA1	−35.16 (45.4)^{c}	41.85 (39.3)	36.00	−30.78 (37.9)^{c}	35.99 (33.0)	37.50	−32.94 (41.8)^{c}	38.89 (36.3)	36.76

PPE, percentage prediction error (known − estimated)/known × 100; |PPE|, absolute percentage prediction error; 20%, percentage of individuals whose estimated body masses fall within ±20% of known mass. Bold numbers indicate the analyses that achieved |PPE|s below 19% and estimated more than 50% of the sample within ±20% of known mass.

[a] Directional differences (positive values indicate underestimation, negative values indicate overestimation).

[b] Absolute differences.

[c] Significance at $P = 0.01$.

TABLE 6. Difference between known and estimated mass, Kappelman (1996) LSR-based equations

	Females ($n = 125$)			Males ($n = 128$)			Combined ($n = 253$)		
Variable	PPE, mean^{a} (SD)	|PPE|, mean^{b} (SD)	20% (%)	PPE, mean^{a} (SD)	|PPE|, mean^{b} (SD)	20% (%)	PPE, mean^{a} (SD)	|PPE|, mean^{b} (SD)	20% (%)
HORB	24.37 (29.9)^{c}	33.39 (19.2)	28.80	34.88 (23.5)^{c}	38.05 (17.9)	21.88	29.69 (27.3)^{c}	35.75 (18.9)	25.30
ORBA3	41.71 (18.9)^{c}	42.80 (16.2)	8.00	45.09 (14.5)^{c}	45.18 (14.2)	6.25	43.40 (16.9)^{c}	43.99 (15.3)	7.11

PPE, percentage prediction error (known − estimated)/known × 100; |PPE|, absolute percentage prediction error; 20%, percentage of individuals whose estimated body masses fall within ±20% of known mass. Bold numbers indicate the variables that achieved |PPE|s below 19% and estimated more than 50% of the sample within ±20% of known mass.

[a] Directional differences (positive values indicate underestimation, negative values indicate overestimation).

[b] Absolute differences.

[c] Significance at $P = 0.01$.

20% criteria. Most of the variables failed to meet the criteria for prediction suitability and overestimated mass significantly ($P = 0.01$). The only variable that met both criteria for acceptance was biporionic breadth and this only occurred in the male sample. Two other variables (biorbital breadth and foramen magnum length) estimated more than 50% of some samples within ±20% of known mass, but failed to meet the |PPE| criterion. The equations for foramen magnum breadth and orbital height were notably poor predictors of mass. In all three test groups (males, females, combined-sex), these equations resulted in |PPE|s over 36% and estimated no more than 28% of the individuals within ±20% of known mass.

The variables that were identified as the best predictors of mass in Aiello and Wood's (1994) study did not perform well in the present study. In particular, orbital area (ORBA1), which showed the lowest errors and highest correlation in Aiello and Wood's (1994) study, returned mean |PPE|s over 28% and failed to estimate more than 35% of any group within ±20% of known mass. The equation for orbital height, which Aiello and Wood (1994:424) considered the "preferred cranial predictor" for large-bodied hominines also returned high errors (>34%) and estimated few individuals (<28%) within ±20% of known mass.

Kappelman's (1996) equations

Kappelman's (1996) two equations also did not predict mass well in the present study. Both equations significantly underestimated mass, and neither met the criteria for acceptance as reliable predictors in any of the three samples (males, females, combined-sexes). The CAD-derived orbital area equation performed particularly poorly, returning |PPE|s in excess of 42% and failing to estimate more than 8% of any sample within ±20% of known mass.

Kappelman's (1996) equations also did not conform to expectations in terms of their performance relative to one another. In Kappelman's (1996) study, orbital area predicted mass more accurately than orbital height in the hominoid sample. In the present study, the equation for orbital height performed consistently and significantly better than the equation for orbital area.

Spocter and Manger (2007) equations

As with the equations of Aiello and Wood (1994) and Kappelman (1996), Spocter and Manger's (2007) LSR equations failed to predict mass reliably in the present study. Most of their equations underestimated mass, and all resulted in estimates that were significantly different from known mass ($P = 0.01$). None met the second criterion for acceptance (50% of individuals within ±20% of known mass) and several (orbital height, biorbital breadth, orbital area as an ellipse, and foramen magnum area as an ellipse) failed to estimate any individuals within ±20% of their known mass. The equations for orbital and foramen magnum area as ellipses produced extremely large prediction errors (>15,000%). It seems

TABLE 7. Difference between known and estimated mass, Spocter and Manger (2007) LSR-based equations

	Females ($n = 125$)			Males ($n = 128$)			Combined ($n = 253$)		
Variable	PPE, mean[a] (SD)	|PPE|, mean[b] (SD)	20% (%)	PPE, mean[a] (SD)	|PPE|, mean[b] (SD)	20% (%)	PPE, mean[a] (SD)	|PPE|, mean[b] (SD)	20% (%)
BORB	43.27 (17.9)	44.07 (15.8)	8.00	43.16 (15.0)	43.24 (14.8)	9.38	43.21 (16.5)[c]	43.65 (15.3)	8.70
HORB	75.96 (9.5)	75.96 (9.5)	0.00	79.30 (7.5)	79.30 (7.5)	0.00	77.65 (8.7)[c]	77.65 (8.7)	0.00
BIOR	73.63 (7.5)	73.63 (7.5)	0.00	73.74 (6.3)	73.74 (6.3)	0.00	73.69 (6.9)[c]	73.69 (6.9)	0.00
BPOR	−40.98 (44.8)	46.74 (38.7)	25.60	−57.81 (37.3)	58.21 (36.7)	15.63	−49.49 (42.0)[c]	52.54 (38.1)	20.55
LFM	28.96 (25.7)	34.78 (16.9)	19.2	29.16 (24.1)	33.11 (18.3)	23.44	29.06 (24.9)[c]	33.94 (17.6)	21.34
BFM	8.98 (35.8)	29.92 (21.5)	42.40	10.74 (31.9)	26.81 (20.4)	45.31	9.87 (33.8)[c]	28.35 (21.0)	43.87
ORBA1	61.58 (12.2)	61.58 (12.2)	0.08	63.92 (9.5)	63.92 (9.5)	0.00	62.76 (11.0)[c]	62.76 (11.0)	0.40
ORBA2	−21542.09 (9526.9)	21542.09 (9526.9)	0.00	−22826.97 (9304.4)	22862.97 (9304.4)	0.00	−22210.36 (9419.5)[c]	22210.36 (9419.5)	0.00
FMA1	16.22 (29.9)	28.20 (19.0)	38.40	17.14 (26.4)	26.86 (16.2)	39.06	16.69 (28.1)[c]	27.52 (17.6)	38.74
FMA2	−13924.50 (8485.9)	13924.50 (8485.9)	0.00	−16678.61 (10561.2)	16678.61 (10561.2)	0.00	−15317.88 (9672.1)[c]	15317.88 (9672.1)	0.00

PPE, percentage prediction error (known − estimated)/known × 100; |PPE|, absolute percentage prediction error; 20%, percentage of individuals whose estimated body masses fall within ±20% of known mass. Bold numbers indicate the variables that achieved |PPE|s below 19% and estimated more than 50% of the sample within ±20% of known mass.

[a] Directional differences (positive values indicate underestimation, negative values indicate overestimation).

[b] Absolute differences.

[c] Significance at $P = 0.01$.

likely that a methodological flaw is responsible for these results. However, repeated discussions (Spocter, pers. comm.) failed to identify the source of the error.

Our results also contrast with those obtained by Spocter and Manger (2007) in terms of the relative performance of their LSR equations. The variables identified as the overall best estimators by Spocter and Manger (2007)—orbital area and orbital area as an ellipse—performed very poorly in the present study.

Interstudy comparison of equations

Orbital height and orbital area are the only variables for which Aiello and Wood (1994), Kappelman (1996), and Spocter and Manger (2007) all provide equations. Of the three for orbital height, Kappelman's (1996) equation resulted in the lowest mean |PPE| and the largest number of individuals estimated within ±20% of their known mass. Aiello and Wood's (1994) equation performed the next best, while Spocter and Manger's (2007) equation performed the most poorly. Kappelman's (1996) orbital height equation also resulted in the lowest raw difference between predicted and known mean mass. Kappelman (1996) and Aiello and Wood's (1994) equations returned similar mean predicted masses (48.7 and 50.2 kg, respectively), and their confidence intervals overlapped considerably. Spocter and Manger's (2007) equation resulted in a very different, and much lower, mean predicted mass (15.9 kg) and the confidence interval did not overlap with those from the other two equations.

With respect to orbital area, Aiello and Wood's (1994) equation predicted mass more accurately than either of Spocter and Manger's (2007) orbital area equations, or Kappelman's (1996) CAD-derived equation. Although neither estimated mass very accurately, Kappelman's (1996) equation performed better than Spocter and Manger's (2007) equation for this variable. Raw predicted masses showed a similar pattern, with Aiello and Wood's (1994) equation resulting in a lower mean difference than the other two equations for orbital area. Mean predicted masses varied considerably between studies and 95% confidence intervals did not overlap for any of the three equations.

It was possible to compare a further six equations between Aiello and Wood's (1994) and Spocter and Manger's (2007) studies. Overall, Aiello and Wood's (1994) equations had lower |PPE|s and estimated more individuals within ±20% of their known mass than Spocter and Manger's (2007) equations. The exceptions to this were the equations for foramen magnum breadth and foramen magnum area (FMA1). Spocter and Manger's (2007) equations for these variables returned lower |PPE|s and estimated more individuals within ±20% of their known mass than those provided by Aiello and Wood (1994). However, the |PPE|s for Spocter and Manger's (2007) equations were still high (>26%) and neither equation estimated more than 45% of any sample within ±20% of known mass. Both Aiello and Wood's (1994) and Spocter and Manger's (2007) equations tended to underestimate known mass. However, the directional differences were not consistent between the two sets of equations. Aiello and Wood's (1994) equations for biporionic breadth foramen magnum length, foramen magnum breadth, and foramen magnum area (FMA1) overestimated mass, while Spocter and Manger's (2007) equations for the same variables underestimated it.

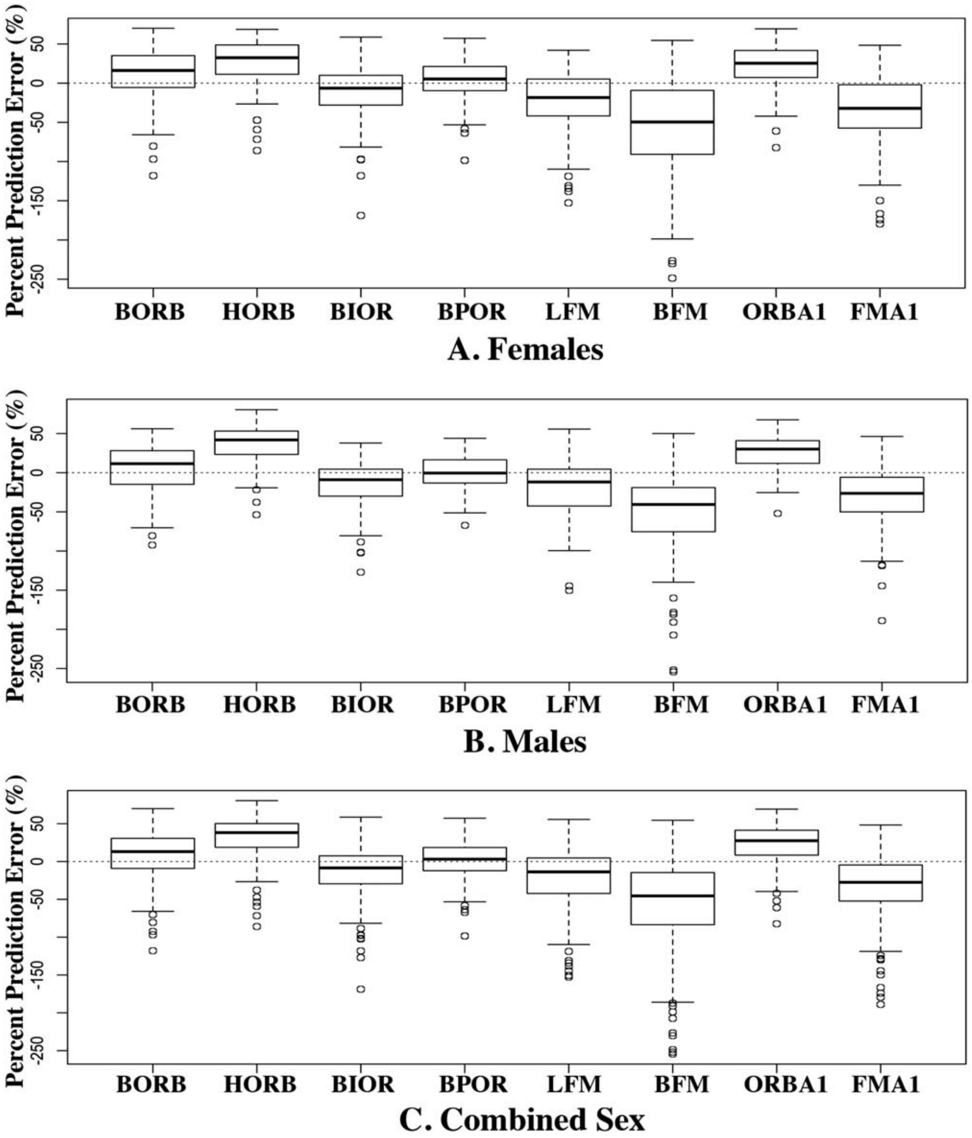
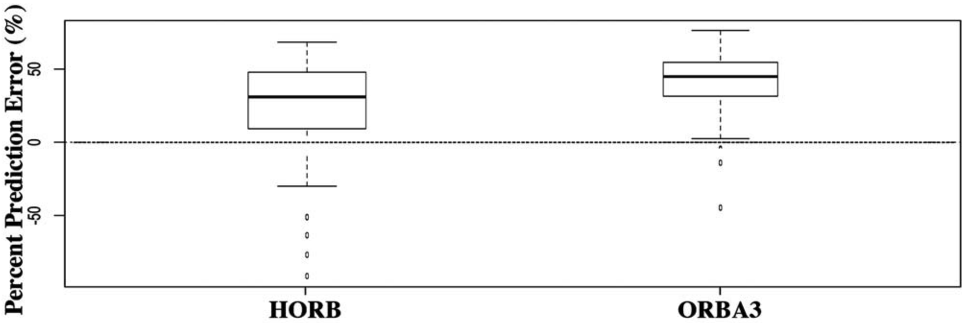


Fig. 2. Box plots of percentage of prediction error (PPE) between known and estimated masses for Aiello and Wood (1994) LSR-based equations: (**a**) females ($n = 125$), (**b**) males ($n = 128$), and (**c**) combined sex ($n = 253$). Solid line = median, upper, and lower box margins = 75th and 25th percentiles, respectively, whiskers = limits of data still within 1.5 interquartile range (IQR) of Q1/Q3.

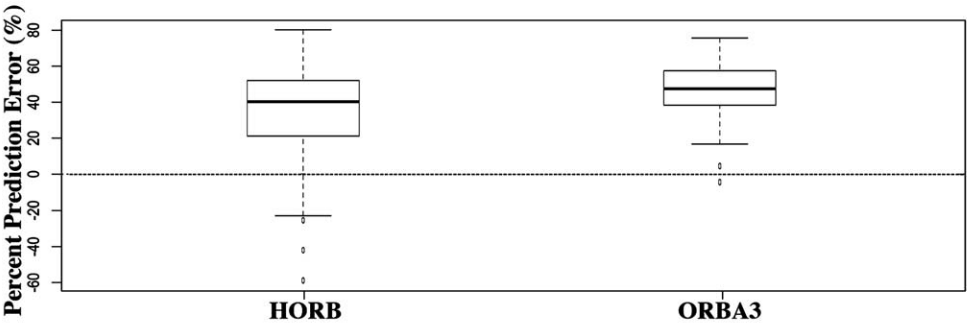
Conversely, Aiello and Wood's (1994) biporionic breadth equation underestimated mass, while Spocter and Manger's (2007) equivalent equation overestimated it. As noted for orbital height and area, Aiello and Wood's (1994) and Spocter and Manger's (2007) equations for the same variables frequently produced very different predicted mean masses. The most extreme example of this was with bi-orbital breadth, which resulted in a mean body mass estimate of 81.7 kg using Aiello and Wood's equation, but just 18.97 kg using Spocter and Manger's (2007) equation. Again, 95% confidence intervals for the mean predicted masses rarely overlapped between the two sets of equations.

DISCUSSION

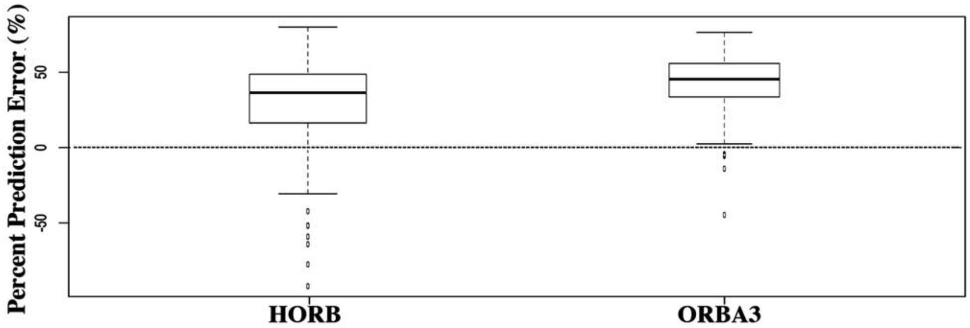
The results of the present study suggest that the existing equations for predicting the body masses of fossil hominins from cranial variables are problematic. As previously noted, Aiello and Wood (1994) found orbital area, orbital height, and biporionic breadth to be good predictors of mass. Kappelman (1996) also considered orbital height and orbital area reliable, while Spocter and Manger (2007) found foramen magnum area, foramen magnum area as an ellipse, biorbital breadth and biporionic breadth useful, with orbital height and area also performing well. However, of the 34 equations tested here, Aiello and Wood's (1994) LSR regression equation for biporionic breadth in males was the only one that resulted in estimates that met both the criteria for acceptance. For the remaining equations, absolute prediction errors exceeded 19% and the number of individuals estimated within ±20% of their known mass was rarely above 50%. In addition, the equations varied markedly in terms of the value of the mass estimated from the same variable, the masses estimated by different variables, and in relation to which variables estimated mass most accurately. Lastly, claims about the advantages of



A. Females



B. Males



C. Combined Sex

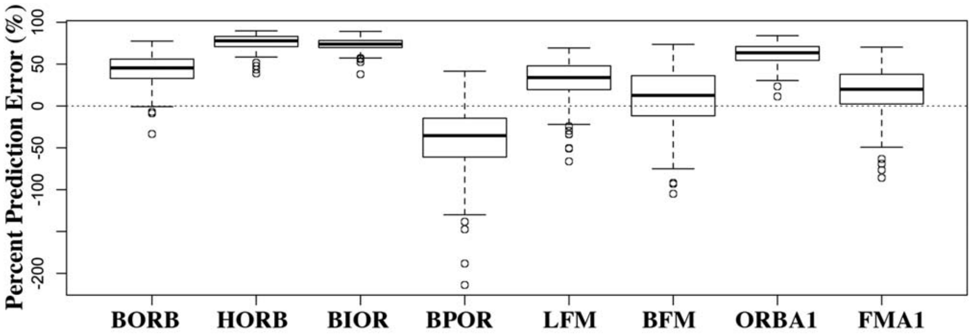
Fig. 3. Box plots of percentage of prediction error (PPE) between known and estimated masses for Kappelman (1996) LSR-based equations: (**a**) females ($n = 125$), (**b**) males ($n = 128$) and (**c**) combined sex ($n = 253$). Solid line = median, upper, and lower box margins = 75th and 25th percentiles, respectively, whiskers = limits of data still within 1.5 interquartile range (IQR) of Q1/Q3.

RMA over LSR-based analyses were not borne out: the RMA-based equations performed worse than all the LSR-based equations.

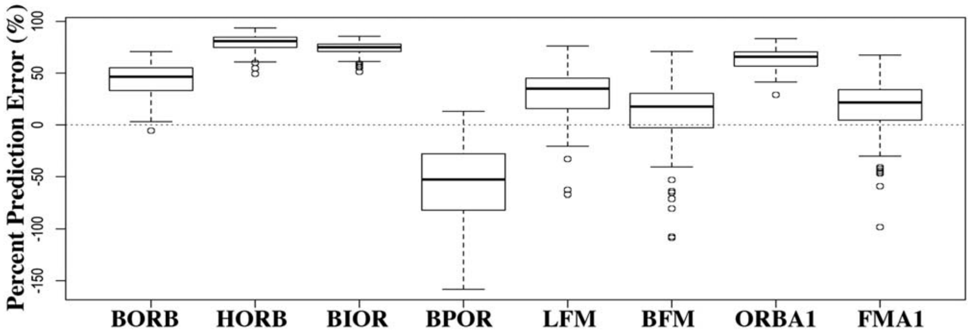
Some of these results were more unexpected than others. The poor performance of the equations for orbital area was particularly surprising. Orbital area was identified as a good predictor of body mass by all three previous studies (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007). It has also been shown to scale the same way in humans and non-human primates (Kappelman, 1996) and has been argued to be "the single best predictor" of body mass in hominins (Churchill et al., 2012:322). However, in the current study, this measurement failed to meet either criterion for acceptance in any sample group, regardless of how it was calculated.

The relative performance of the different area equations was also surprising. For both orbital area and foramen magnum area the breadth × height equations performed better than either the ellipse or CAD-derived equations. If these features were meaningfully related to body mass, then the more accurate calculations should have performed better (Kappelman, 1996). However, this was not the case and neither the ellipse equation nor the CAD-derived equation achieved acceptable rates of accuracy in the sample group.

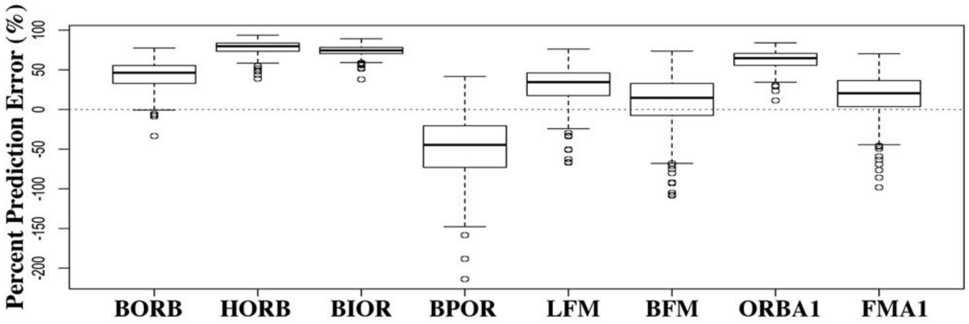
Several potential problems need to be considered before the results of this study can be accepted. The first relates to estimating the body mass of a single individual within a species from an interspecies sample. As noted previously, all the equations were derived from sex/species means but are used regularly to estimate the body mass of individual fossil specimens. Despite this, it is possible that within-species scaling of cranial variables with body mass in *Homo sapiens* is so different from among-species scaling of the same variables in non-human primates that the interspecific equations are incapable of accurately predicting body mass of an individual human. There are two reasons we believe this is an unlikely explanation for the current results. First,



A. Females



B. Males



C. Combined Sex

Fig. 4. Box plots of percentage of prediction error (PPE) between known and estimated masses for Spocter and Manger (2007) LSR-based equations: (**a**) females ($n = 125$), (**b**) males ($n = 128$), and (**c**) combined sex ($n = 253$). Solid line = median, upper, and lower box margins = 75th and 25th percentiles, respectively, whiskers = limits of data still within 1.5 interquartile range (IQR) of Q1/Q3.

orbital area has been found to scale in the same way with body mass in humans as it does in other primates (Kappelman, 1996). In the current study, three different equations for this variable were tested and none resulted in good predictive ability. While such effects may still be responsible for errors in other variables, this suggests that the equations' poor performance is not solely a consequence of the variables scaling differently within modern humans than among non-human primates.

The second reason for suspecting that the poor performance of the equations is not due to scaling differences is that we carried out a supplementary analysis, and its results were not consistent with this explanation. In the analysis in question, we generated body mass estimates from the means of 50 sets of ten individuals each (randomly selected with replacement) and compared these values with the known mean masses. Aiello and Wood's (1994) equations were used for this test as they yielded the most accurate estimates in the initial analyses. If scaling factors were the issue, the prediction errors were expected to be consistently lower when mean data were used. Contrary to this, prediction errors were lower in some cases but higher in others (Supporting Information Table S4).

A second potentially confounding factor is the inclusion of very light and very heavy individuals in the test sample. Because it consisted of modern Europeans, the current sample included individuals with a wide range of BMIs (14–45). Although the mean BMI (25.46) corresponds closely with the current national average (24.6) for Switzerland (SFSO, 2012), this BMI range likely differs significantly from that of fossil hominins and early modern human populations, particularly in the "overweight" category. Consequently, it is possible that

TABLE 8. Mean body mass differences and confidence intervals for the combined sex sample ($n = 253$)*, LSR-based equations*

Variable	Source	Mean predicted BM (kg)	Difference from known mean[a] (kg)	95% CI for predicted mass
BORB	Aiello and Wood (1994)	66.48	9.14	64.09–68.87
	Spocter and Manger (2007)	40.84	34.78	39.80–41.89
HORB	Aiello and Wood (1994)	48.67	26.95	46.97–50.36
	Kappelman (1996)	50.17	25.45	48.41–51.92
	Spocter and Manger (2007)	15.95	59.67	15.39–16.51
BIOR	Aiello and Wood (1994)	81.71	6.09	79.42–84.00
	Spocter and Manger (2007)	18.97	56.65	18.56–19.39
BPOR	Aiello and Wood (1994)	71.80	3.82	70.00–73.61
	Spocter and Manger (2007)	108.96	33.34	105.45–112.48
LFM	Aiello and Wood (1994)	87.14	11.52	84.58–89.70
	Spocter and Manger (2007)	51.26	24.36	49.32–53.19
BFM	Aiello and Wood (1994)	110.95	35.33	106.52–115.38
	Spocter and Manger (2007)	64.95	10.67	62.34–67.57
ORBA1	Aiello and Wood (1994)	54.62	21.00	53.15–56.09
	Spocter and Manger (2007)	26.70	48.92	26.08–27.33
ORBA3	Kappelman (1996)	40.52	35.10	39.54–41.50
FMA1	Aiello and Wood (1994)	95.76	20.14	92.81–98.72
	Spocter and Manger (2007)	60.13	15.49	58.00–62.26

[a] Mean known body mass for full sample ($n = 253$) is 75.62 kg.

the equations performed poorly in the present study because the range of variation greatly exceeded that expected by the method. To test this possibility, we ran an additional set of analyses on a subsample of individuals ($n = 116$) with BMIs in what the World Health Organization considers to be the "normal" range (18.5–25) (WHO, 2000). Supporting Information Tables S5 and S6 provide the sample and variable summaries for the BMI-restricted groups, while Supporting Information Tables S7–S9 summarize the results for the three sets of equations. Restricting the sample to a normal BMI range did not consistently improve accuracy. Several of Aiello and Wood's (1994) and Spocter and Manger's (2007) equations returned lower |PPE|s and estimated more individuals within ±20% of known mass in the BMI-restricted samples. However, other equations produced higher |PPE|s and estimated fewer individuals within ±20% of known mass, and most equations still did not meet both criteria for acceptance. Both of Kappelman's (1996) equations returned lower |PPE|s and estimated more individuals within ±20% of known mass, but still failed to meet the criteria for acceptance. Thus, the inclusion of very light and very heavy individuals also does not explain the poor performance of the equations.

The inclusion of older individuals in the sample is also potentially problematic. Body mass can change significantly over the course of a lifetime, and as past populations are less likely to have lived into very old age (Robson and Wood, 2008), it is possible that the presence of older individuals negatively affected the results (Ruff, pers. comm.). To evaluate this, we ran another set of analyses using only individuals between 18 and 60 years of age. This reduced the test sample to 87 females and 99 males (total $n = 186$). Again, as they produced the most reliable results overall, Aiello and Wood's (1994) equations were used for this test. The results show that constraining the sample to a more "realistic" age range for fossil hominins had little effect on accuracy (see Supporting Information Table S10). Prediction errors were variably higher or lower than those for the full sample, but none was significantly different ($P = 0.01$). This suggests that the poor performance of the equations in the present study was not due to the inclusion of very old individuals in the sample either.

It appears, then, that the poor performance of the equations is not due to shortcomings in the design of our study, but to some other factor or combination of factors.

What else might be driving the poor performance of the equations? One possibility relates to the lack of a functional relationship between cranial morphology and body mass. Several authors have argued that body mass will be better predicted by skeletal features that are functionally related to bearing weight—e.g., the proximal femur (Jungers, 1988; Ruff, 1991, 1994; McHenry, 1992). Based on this, cranial variables would not be expected to predict mass very well (Hylander, 1985). However, the existence of a functional relationship is not essential for a trait to be useful for prediction (Smith, 2002). Indeed, the very *lack* of such a relationship has been argued to be the primary advantage of postcranial equations that employ stature and bi-iliac breadth to estimate body mass compared to postcranial equations that utilize femur head breadth (Ruff et al., 1997, 2005, 2012). Thus, it seems unlikely that the absence of a functional relationship between the cranial variables and body mass explains the poor performance of the majority of the cranial equations tested in the present study.

The quality of the data is a more likely source of error. As individual body masses are rarely available for wild-caught non-human primates (Kappelman, 1996), the studies in question all employed measurements taken from one set of specimens and body mass data taken from the literature (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007). While this allows more species to be included and larger sample sizes, the use of unassociated data necessarily reduces the ability to characterize the relationship between cranial morphology and mass. Consequently, the accuracy and reliability of the original data used to generate the predictive equations are not assured (Ericksen, 1982; Komar and Grivas, 2008).

Sample size may be a much bigger problem. As noted previously, sample sizes in the three studies were very small. Specifically, most of the non-human primate taxa in Aiello and Wood's (1994) study were represented by 10 individuals and the human sample consisted of only 12 males and 12 females. Kappelman (1996) used between 5 and 30 individuals for the non-human primate species and 32 individuals for the human sample. Spocter and Manger (2007) used a larger sample of modern humans (90 males and 90 females), but their non-human primate samples were represented by only two or three individuals. In fact, because all three studies used mean data to generate their predictive equations, sample sizes were effectively reduced to the number of sex/species data points. For the hominoid-based equations this resulted in sample sizes of 5, 12, and 18 in Spocter and Manger's (2007), Aiello and Wood's (1994), and Kappelman's (1996) analyses, respectively. However, such small samples significantly increase the probability of nonrandom sampling (Ruff, 2003) and cannot be considered sufficient for statistically robust interpretations (Dupont and Plummer, 1998; Smith, 2002). In addition, small sample sizes can produce artificially high correlation coefficients and consequently, misleadingly reliable results. This is particularly true of RMA regression analyses (Legendre, 1998), although with high r^2 values, the slopes of RMA and LSR are similar and LSR methods are likely to perform as poorly as RMA methods when sample sizes are limited. Consequently, the largest contributor to the equations' poor performance may be the small samples of the reference material used to generate them.

The results of the present study have several implications for human evolutionary research. The most obvious of these is the need to prioritize the task of improving the estimation of fossil hominin body mass from cranial variables. Given the challenges of attributing postcranial material to specific taxa (Aiello and Wood, 1994) and the fact that the fossil record continues to be weighted heavily toward cranial material (e.g., Ji et al., 2013), it seems premature to discourage the estimation of body mass from such material without at least attempting to correct existing problems. One route would be to develop new predictive equations using larger sample sizes. Although challenging (see Wood and Collard, 1996), incorporating larger groups of non-human primates with associated skeletal dimensions and body mass data into these analyses would be ideal. Alternatively, for fossil hominins, particularly those in *Homo*, it may be more appropriate to employ an exclusively modern human sample for the new analyses. Deriving sex-specific equations from such a sample may also improve predictive ability. Comparisons with postcranial material in the same associated human sample might lead to the identification of cranial variables that are more effective. Finally, alternate statistical approaches could also be considered. In particular, Uhl et al. (2013) demonstrate the utility of using Bayesian and maximum likelihood methods for estimating body mass and recommend using R statistics to explicitly consider differences in size and scaling between modern and fossil samples. Konigsberg and Frankenberg (2013) expand on this concept and provide useful guidance for employing Bayesian methods for a variety of questions in biological anthropology.

A second implication relates to the choice of RMA-based equations versus LSR-based equations. As noted previously, there is currently no consensus regarding the line-fitting method that should be used when developing equations to predict fossil hominin body mass from skeletal variables. Some researchers argue that LSR should be used, while others contend that RMA regression is more appropriate. Although the underlying problem here may have more to do with the sample size of the original reference groups than anything else (see above), the results of the present study suggest that LSR may be the more accurate method for predicting the body mass of fossil hominins. To confirm this, reanalyses of both methods using large and representative samples are clearly needed.

A third implication of our results concerns the interpretation of the body mass estimates that have already been generated with the cranial equations. If prediction errors are high using a sample of individuals whose species is represented in the reference sample, it seems likely that they will be at least as high for specimens whose species are not represented, as is the case for the fossil hominins. Thus, our results suggest that most of the body mass estimates for the fossil hominins that have been derived from cranial equations (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007; Churchill et al., 2012) should be treated only as very rough "ball park" figures. Given that most of the equations failed to accurately estimate more than 50% of the specimens in our sample within ±20% of their known mass, it would seem reasonable to allow for the possibility that the body mass estimates for many fossil hominins generated with existing equations are at least 20% too low or too high.

CONCLUSIONS

Overall, existing equations for estimating body mass from cranial variables produced high rates of error in a sample of modern humans. Despite methodological similarities, estimates between the studies reviewed varied considerably, as did the relative performance of different equations for the same variables. In particular, variables that had previously been identified as good predictors of body mass in hominoids, were not the most reliable in the human sample. In addition, RMA regression methods were not found to be more appropriate than LSR methods for predicting body mass from cranial variables. Problems with the size and composition of the original samples may be largely responsible for the failure of the equations to predict mass adequately. Consequently, further analyses involving larger samples and careful consideration of reference and target groups are warranted. New statistical approaches may also improve predictive ability. With refinements such as these, it may still be possible to increase the accuracy of body mass estimates in fossil hominins using cranial variables.

ACKNOWLEDGMENTS

Many thanks to Michael Thali, Wolf Schweitzer, and the rest of team at the Institute for Forensic Medicine in Zurich, Switzerland, for granting access to the raw CT scans and for their expertise and assistance during the data collection. Thanks also go to the members of the Human Evolutionary Studies Program for helpful observations and to Kimberly Rondeau and Isabella Ghement for statistical assistance. Final thanks go to Chris Ruff and two anonymous reviewers for constructive improvements to this article.

LITERATURE CITED

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

Aiello LC, Key C. 2002. Energetic consequences of being a *Homo erectus* female. Am J Hum Biol 14:551–565.

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

Calder WA. 1984. Size, function, and life history. Cambridge: Harvard University Press.

Cavalcanti M, Rocha S, Vannier M. 2004. Craniofacial measurements based on 3D-CT volume rendering: implications for clinical applications. Dentomaxillofac Radiol 33:170–176.

Barrickman NL. 2008. Evolutionary relationship between life history and brain growth in anthropoid primates. Unpublished PhD thesis, Duke University, Durham, NC.

Churchill SE, Berger LR, Hartstone-Rose A, Zondo BH. 2012. Body size in African Middle Pleistocene *Homo*. In: Reynolds SC, Gallagher A, editors. African genesis: perspectives on hominin evolution. Cambridge: Cambridge University Press.

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 13:307–344.

Damuth J, MacFadden B. 1990. Body size in mammalian paleobiology. Cambridge, UK: Cambridge University Press.

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 Forensic Sci 5:1107–1114.

DeSilva JM, Lesnik JJ. 2008. Brain size at birth throughout human evolution: a new method for estimating neonatal brain size in hominins. J Hum Evol 55:1064–1074.

Dupont WD, Plummer WD Jr. 1998. Power and sample size calculations for studies involving linear regression. Control Clin Trials 19:589–601.

Ericksen MF. 1982. How "representative" is the Terry Collection? Evidence from the proximal femur. Am J Phys Anthropol 59:345–350.

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

Hartwig-Scherer S, Martin RD. 1992. Allometry and prediction in hominoids: a solution to the problem of intervening variables. Am J Phys Anthropol 88:37–57.

Hylander WL. 1985. Mandibular function and biomechanical stress and scaling. Am Zool 25:315–330.

Ji, XP, Jablonski, NG, Su, DF, Deng, CL, Flynn, LJ, You, YS, Kelley, J. 2013. Juvenile hominoid cranium from the terminal Miocene of Yunnan, China. Chin Sci Bull 1–9.

Jungers WL. 1988. Relative joint size and hominoid locomotor adaptations with implications for the evolution of hominid bipedalism. J Hum Evol 17:247–265.

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

Komar DA, Grivas CR. 2008. Manufactured populations: what do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum documented collection. Am J Phys Anthropol 137:224–233.

Konigsberg LW, Frankenberg SR. 2013. Bayes in biological anthropology. Am J Phys Anthropol 152:153–184.

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, D. 2001. Primates from Rudabanya: allocation of specimens to individuals, sex and age categories. J Hum Evol 40:17–39.

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

Legendre P. 1998. Model II regression user's guide, R edition. R Vignette.

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: 405.e1–405.e6.

Lopes PML, Moreira CR, Perrella A, Antunes JL, Cavalcanti, MGP. 2008. 3-D volume rendering maxillofacial analysis of angular measurements by multislice CT. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 105:224–230.

Martin R. 1928. Lehrbuch der anthropologie in systematischer darstellung mit besonderer berücksichtigung der anthropologischen methoden für studierende ärtze und forschungsreisende: G. Fischer.

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

McHenry HM, Coffing K. 2000. *Australopithecus* to *Homo*: transformations in body and mind. Ann Rev Anth 29:125–114.

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 147:264–279.

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

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

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

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

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

Ruff CB. 2003. Long bone articular and diaphyseal structure in Old World monkeys and apes. II: estimation of body mass. Am J Phys Anthropol 120:16–37.

Ruff CB. 2005. Mechanical determinants of bone form: insights from skeletal remains. J Musculoskelet Neuronal Interact 5: 202–212.

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 86:397–413.

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

Ruff C, Niskanen M, Junno JA, 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 48:381–392.

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

Sciulli PW, Blatt SH. 2008. Evaluation of juvenile stature and body mass prediction. Am J Phys Anthropol 136:387–393.

Smith RJ. 1996. Biology and body size in human evolution: statistical inference misapplied Curr Anthropol 37:451–481.

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

Smith RJ. 2009. Use and misuse of the reduced major axis for line-fitting. Am J Phys Anthropol 140:476–486.

Smith RJ, Jungers WL. 1997. Body mass in comparative primatology. J Hum Evol 32:523–559.

Smyth AM, Viner MD, Conlogue GJ, Blyth T. 2012. An evaluation of the use of modern medical imaging techniques for the determination of biological sex from craniometric measurements. Proc Am Acad Forensic Sci 28:361.

Sokal R, Rohlf F. 1995. Biometry. New York, NY: WH Freeman & Co.

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

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

Thali, MJ, Jackowski, C, Oesterhelweg, L, Ross, SG, Dirnhofer, R. 2007. Virtopsy—the Swiss virtual autopsy approach. Leg Med 9:100–104.

Uhl, NM, Rainwater, CW, Konigsberg, LW. 2013. Testing for size and allometric differences in fossil hominin body mass estimation. Am J Phys Anthropol 151:215–229.

Wood B, Collard M. 1999. The human genus. Science 284: 65–71.

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

World Health Organization. 2000. Obesity: preventing and managing the global epidemic. Report on a WHO Consultation Technical Report Series, No. 894. Geneva: World Health Organization.

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