Human Body Mass Estimation: A Comparison of "Morphometric" and "Mechanical" Methods

Benjamin M. Auerbach* and Christopher B. Ruff

Center for Functional Anatomy and Evolution, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

ABSTRACT   In the past, body mass was reconstructed from hominin skeletal remains using both "mechanical" methods which rely on the support of body mass by weight-bearing skeletal elements, and "morphometric" methods which reconstruct body mass through direct assessment of body size and shape. A previous comparison of two such techniques, using femoral head breadth (mechanical) and stature and bi-iliac breadth (morphometric), indicated a good general correspondence between them (Ruff et al. [1997] Nature *387:*173–176). However, the two techniques were never systematically compared across a large group of modern humans of diverse body form. This study incorporates skeletal measures taken from 1,173 Holocene adult individuals, representing diverse geographic origins, body sizes, and body shapes. Femoral head breadth, bi-iliac breadth (after pelvic rearticulation), and long bone lengths were measured on each individual. Statures were estimated from long bone lengths using appropriate reference samples. Body masses were calculated using three available femoral head breadth (FH) formulae and the stature/bi-iliac breadth (STBIB) formula, and compared. All methods yielded similar results. Correlations between FH estimates and STBIB estimates are 0.74–0.81. Slight differences in results between the three FH estimates can be attributed to sampling differences in the original reference samples, and in particular, the body-size ranges included in those samples. There is no evidence for systematic differences in results due to differences in body proportions. Since the STBIB method was validated on other samples, and the FH methods produced similar estimates, this argues that either may be applied to skeletal remains with some confidence. Am J Phys Anthropol 125:331–342, 2004.   © 2004 Wiley-Liss, Inc.

Despite inherent limitations (Smith, 1996), body mass estimation from human skeletal remains will continue to play an important role in paleontological and archaeological studies, for a variety of reasons (Ruff, 2000a). While cranial dimensions can be used to predict body mass (Aiello and Wood, 1994; Kappelman, 1996), it is generally agreed that postcranial features, which have a more direct relationship to overall body size, produce the most accurate estimates (Jungers, 1988; Ruff et al., 1989; Aiello and Wood, 1994; Hartwig-Scherer, 1994; Delson et al., 2000; Porter, 2002).

Two general approaches to body mass estimation using postcrania have been employed: "mechanical" methods, which rely on the functional association between a weight-bearing skeletal element and body mass, and "morphometric" methods, which attempt to directly reconstruct body size and/or shape from preserved elements (Ruff, 2002). Mechanical methods may be further subdivided into those that employ articular surface dimensions, and those that use diaphyseal breadths or cross-sectional dimensions. Articular surface dimensions have the advantage of being influenced less by differences in activity level or muscular loadings during life than diaphyseal dimensions (Ruff, 1988; Trinkaus et al.,

1994; Lieberman et al., 2001). Femoral head breadth is particularly valuable in this regard because it is frequently available in human archaeological and paleontological specimens, is an easily taken and highly reproducible measurement, and was included in several studies of modern humans with known body masses, producing three sets of body mass estimation formulae (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995).

Morphometric methods for estimating body mass based on stature reconstruction alone have also been developed (Mathers and Henneberg, 1995; Porter,

1995). These, however, do not account for the great variation in body shape, i.e., body breadth relative to length, observable in modern and earlier humans (Ruff, 1994, 2002). A morphometric method combining stature and body (bi-iliac, or maximum pelvic) breadth, based on a worldwide sampling of modern humans of diverse body shape, was shown to provide relatively accurate estimates in both "normal" and highly athletic modern humans (Ruff, 2000a).

It is of interest to compare body mass estimates obtained using the mechanically based femoral head methods and the morphometric stature/bi-iliac breadth method. Similarity in results between methods would increase confidence in the validity of each, as each technique is based on very different skeletal dimensions and underlying rationales. A previous comparison of body mass estimates using these methods did produce reasonably congruent results in a sample of 75 Pleistocene *Homo* specimens (Ruff et al., 1997). However, this sample was necessarily limited in both its size and the relative completeness of material: in many cases, bi-iliac breadth was estimated, and stature estimates were less secure than in more complete specimens whose body proportions and/or relationships to modern reference samples are better known.

The present study greatly expands this comparison to a sample of more than 1,000 relatively complete Holocene human skeletons of diverse geographic origins, body sizes and body shapes, and known pelvic breadths. This much larger sample allows closer examination of possible differences between techniques and the underlying causes of these differences. In addition, the three available femoral head methods are individually evaluated for correspondence with the stature/bi-iliac breadth technique, in order to highlight the potential advantages and disadvantages of each method. Based on our results, some general recommendations are made for the use of these methods in estimating body mass from skeletal remains.

METHODS

Samples and measurements

Metric dimensions were taken from the skeletal remains of 1,173 adult Holocene humans (745 males, 428 females), spanning six continents and several dozen populations and nationalities (Table 1). For various comparisons, the total sample was subdivided into six broad or 17 narrower regional groups, as defined in Table 1. Museum sources for all materials are listed in Table 1. Data for two African Pygmies were obtained from the literature (Flower, 1989; van der Broek, 1940); otherwise, all measurements were obtained from original specimens by the authors. All individuals, except for 163 Europeans from two collections (Hamann-Todd first-generation European immigrants and a population from Syracuse, Sicily) are preindustrial, dating from the Late Bronze Age to early Medieval

period in Europe, pre-European contact among Amerindians, and nonindustrial 19th and 20th century populations from other locations.

Femoral head anterior-posterior breadth, maximum long bone lengths, femoral bicondylar length, and bi-iliac pelvic breadths (after rearticulation; see Ruff, 2000b) were taken on each individual. When possible, measurements were taken on both sides and averaged to reduce laterality biases. In tibial maximum length measures (used only in estimating Pygmy statures), the intercondyloid eminences ("spines") and malleolus were included. Measurements were taken by the authors with dial calipers, digital dial calipers linked directly to a notebook computer, and a portable osteometric board. Only individuals with fully fused epiphyses on both the long bones and the pelvis were measured to ensure that primary individual growth had stopped. Any individuals showing pathologies due to age, disease, or trauma were removed from the study. All measures used here were checked by remeasuring a random sample of previously observed skeletal remains while at each collection. The difference between the initial measurement and the remeasurement of each bone metric was less than 1%; thus all measures are considered highly reliable.

Sexing of noncadaveric individuals was accomplished by the method of Phenice (1969), using features of the os coxae. Recorded sexes were used in the few skeletons derived from cadavers. In the instances in which sexing from the pelvis was ambiguous on noncadaveric remains, visual characters were observed on available crania. All individuals for whom sexing remained uncertain were not included in this study.

Stature estimation

Because stature is used in the stature/bi-iliac body mass estimation technique, and because the statures in this study can only be estimated from skeletal remains (with the exception of the few cadaveric samples), it is crucial to carefully consider the limitations involved in such approximations. Recently addressed issues include the choice of line fitting technique, skeletal element(s) used, and choice of reference population (Krogman, 1962; Holliday and Ruff, 1997; Konigsberg et al., 1998; Hens et al., 2000). We chose to use (ordinary) least squares equations of stature regressed on maximum femur length for most of the samples here, because such equations are available for a wide variety of reference samples (Table 2), and have relatively low associated standard errors of estimate. In two cases, we used maximum tibia length, alone or in combination with femoral length, because only the tibia had been measured (Allbrook, 1962) or because the multivariate femur and tibia equation had a significantly smaller SEE than that for the femur alone (Olivier, 1976). Sex-specific formulae were used whenever possible. The only exceptions were these latter two equations, in which only males had been

TABLE 1. Populations used in this study

Broad group	Narrow group (group number)	Population/nationality	N	Source[1]
Andaman Islanders	Andaman Islanders (1)	Great Andamanese, Jarawa	8	AMNH; DC; NHM
African Pygmies	African Pygmies (2)	Region d'lrumu	3	IRSN
		Akkas	1	Flower (1889)
		Efé	1	Van der Broek (1940)
Australian Aborigines	Australian Aborigines (3)	Several populations[2]	24	AMNH; DC; MdH; MNdAE; NMNH; NM; NHM
East Africans	Kenyans and Ugandans (4)	Northeastern Uganda and northwestern Kenya	46	MU KNM
North Africans	Egyptians (5)	Early Dynastic to Middle Kingdom Egyptian (4 sample groups)	66	AMNH; MdH; NM
	Nubians (6)	Nubians and Kerma (2 sample groups)	36	DC; NM
Europeans[3]	Europeans (7)	Austrians (3 sample groups)	71	NM
		Belgians (2 sample groups)	29	IRSN
		English and Scottish (4 sample groups)	77	DC; NHM
		French (2 sample groups)	23	MdH
		Germans (4 sample groups)	73	SfAP
		Italians (3 sample groups)	46	MNdAE
		Hamann-Todd first-generation immigrants (7 sample groups)[4]	163	CMNH
Amerindians	Yamana (8)	Fuegians (Yamana; 1 sample group)	21	MdH; MNdAE
	Arikara (9)	Arikara (3 sample groups)	55	NMNH
	Illinois Hopewell (10)	Illinois Hopewell (2 sample groups)	40	NMNH
	Indian Knoll (11)	Indian Knoll (1 sample group)	67	WOAC
	Alaskan Inuit (12)	Alaskan Inuit (2 sample groups)	108	AMNH; NMNH
	Aleutian Islanders (13)	Aleutian Islanders (4 sample groups)	51	NMNH
	Libben (14)	Libben (1 sample group)	28	KSU
	Pueblo (15)	Pueblo (4 sample groups)	104	MCZ; NMNH
	Santa Cruz Islanders (16)	Santa Cruz Island (1 sample group)	13	NHM
	Utes (17)	Utes (1 sample group)	29	AMNH

[1] AMNH, American Museum of Natural History, New York City, NY; CMNH, Cleveland Museum of Natural History, Cleveland, OH; DC, Duckworth Osteological Collection, Cambridge, United Kingdom; IRSN, Institut Royal des Sciences Naturelles de Belgique, Brussels, Belgium; KNM, Kenya National Musem, Nairobi, Kenya; KSU, Kent State University (Libben Collection), Kent, OH; MCZ, Museum of Comparative Zoology, Cambridge, MA; MdH, Musée de l'Homme, Paris, France; MNdAE, Muzeo Nationale di Antropologia e Etnologia, Florence, Italy, MU, Makerere University, Kampala, Uganda; NHM, British Museum (Natural History Museum), London, United Kingdom; NM, Naturhistorishes Museum, Vienna, Austria; NMNH, National Museum of Natural History (Smithsonian Institution), Washington, DC; SfAP, Staatssammlung für Anthropologie und Palaeoanatomie, Munich, Germany; WOAC, Webb Osteology and Archaeology Collection, Lexington, KY.
[2] Aboriginal individuals in these collections are not of the same provenances, nor are the original locations from which some individuals came known. We are aware of variation among native Australians from different regions, and that combining individuals in this manner does not reflect these differences. However, for this study, all Australian aborigines were grouped together with the confidence that all individuals tested have similar crural indices, and therefore similar body proportions.
[3] Although populations are here listed within modern country borders, this is not meant to reflect the social and biological history of many of the populations measured.
[4] Individuals were attributed to seven countries of origin based on Todd's records: Austria, Germany, Ireland, Hungary, Italy, Poland, and Russia.

measured. In these cases, we used the male equations on both sexes, with the rationale that differences in linear body proportions between the sexes within a population are likely be less than those between two different populations (Sjøvold, 1990).

The study samples were matched with stature formula reference samples as closely as possible, using population affinities when available (e.g., African Pygmies estimated using the Pygmy formula of Olivier, 1976; East Africans using the East African formula of Allbrook, 1961; and US Midwestern Amerindians estimated using the prehistoric Ohio formula of Sciulli and Giesen, 1993), or in the absence of such affinities, using linear body proportions as a guide. The primary body proportion evaluated in this respect was the crural index, or tibia/femur length. Although this is an imperfect index of linear body proportions (Holliday, 1997; Ruff et al.,

2002), it does give some indication of relative limb to trunk length (Ruff et al., 2002), and thus an appropriate reference sample (Holliday and Ruff, 1997). All stature estimation equations are listed in Table 2 with the groups to which they were applied. Further details of stature techniques employed are included in the Appendix.

Body mass estimation

Four body mass estimates, calculated using previously published equations, were obtained for each individual. All equations are listed in Table 2. They include three femoral head body mass estimation (FH BME) techniques (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995) and a stature/bi-iliac breadth body mass estimation (STBIB BME) method (Ruff, 2000b; Ruff et al., 1997). Two of the formulae are sex-specific the FH BME of Ruff et al.

TABLE 2. Equations

Source	EQUATION	Sample composition	Groups applied to[1]
Stature equations[2]			
Allbrook (1961)	$S = 2.00 \times Tib_m + 91.19$	Nilohamite people, Teso, Kenya	3, 4
Genoves (1967)	$S = 2.26 \times Fem_m + 66.379 - 2.5$ (males)	Cadaveric Mexican Hispanics (early) 20th century	8, 15, 16, 17
	$S = 2.59 \times Fem_m + 49.742 - 2.5$ (females)		
Olivier (1976)	$S = 1.31 \times (Fem_b + Tib_m) + 55.3$	Cameroon Baka Pygmies	1, 2
Sciulli and Giesen (1993)	$S = 2.443 \times Fem_m + 42.805$ (males)	Early Woodland Ohio Valley Native Americans	9, 10, 11, 14
	$S = 2.336 \times Fem_m + 44.253$ (females)		
Trotter and Gleser (1952)	$S = 2.38 \times Fem_m + 61.41$ ("white" males)	Cadaveric United States whites and blacks (early 20th century)	5, 6, 7, 12, 13
	$S = 2.47 \times Fem_m + 54.10$ ("white" females)		
	$S = 2.11 \times Fem_m + 70.35$ ("black" males)		
	$S = 2.28 \times Fem_m + 59.76$ ("black" females)		
Femoral head body mass estimation (FH BME) equations[3]			
Ruff et al. (1991)	$BM = (2.741 \times FH - 54.9) \times .90$ (males)	Based on data for 80 individuals taken from a population from Baltimore, MD	
	$BM = (2.426 \times FH - 35.1) \times .90$ (females)		
	$BM = (2.160 \times FH - 24.8) \times .90$ (combined sex)[4]		
McHenry (1992)	$BM = 2.239 \times FH - 39.9$	Based on 4 sample means of North American males and females, African Pygmies, and Khoisan	
Grine et al. (1995)	$BM = 2.268 \times FH - 36.5$	Based on 10 sex-specific means for samples of large-bodied African Americans, European Americans, and Native Americans	
Stature/bi-iliac breadth body mass estimation (STBIB BME) equations[5]			
Ruff et al. (1997)	$BM = 0.373 \times ST + 3.033 \times LBIB - 82.5$ (males)	Based on 56 sex-sample means for a worldwide sample of modern humans; original data in Ruff (1994)	
	$BM = 0.522 \times ST + 1.809 \times LBIB - 75.5$ (females)		

[1] See Table 1 for narrow-group number references.

[2] S, estimated stature in cm; Fem_m, femoral maximum length (cm); Fem_b, femoral bicondylar length (cm); Tib_m, maximum length of tibia (cm).

[3] BM, body mass in kilograms; FH, femoral head anterior-posterior breadth (mm).

[4] Not used in this study; included for reference in cases of unknown sex.

[5] BM, body mass in kilograms; ST, stature in centimeters; LBIB, living bi-iliac breadth in centimeters ($1.17 \times$ skeletal BIB $- 3$) (Ruff et al., 1997).

(1991), and the STBIB BME of Ruff et al. (1997). The other formulae are for combined sexes. All formulae, except those from Ruff et al. (1991), are based on sex and population means.[1] Composition of the reference samples is given in Table 2.

The FH BME techniques were individually tested against the STBIB BME technique, by population, sex, and combined samples. As no known masses for individuals were incorporated into this study, the STBIB BME was used as a nominal "standard" against which the other techniques could be collectively compared (for overall correspondence between "mechanical" and "morphometric" methods), and to assess individual differences among FH BMEs. We do not mean to imply that the results of the STBIB BME technique could generally be taken as proxies for known body masses, but the STBIB BME is based on the most diverse sample of any of the body mass estimation formulae reviewed in this paper (Ruff, 1994), and was validated on other samples (Ruff, 2000a). Hence, it is a useful baseline for comparing these other methods.

Several types of comparisons between techniques were carried out. Bivariate scatters of each FH BME against the corresponding STBIB BME over the en-

tire pooled sample were generated and used to evaluate overall correspondence between techniques and possible size-related effects. To maintain proportionality over the size range represented in the sample, data were plotted on a logarithmic scale. Reduced major axis (RMA) regressions were fit through each scatter and compared to lines of equivalence. RMA, rather than least squares, was used here due to the uncertainty inherent in both estimates (Smith, 1996).

Average differences between technique results were calculated as both raw (FH BME − STBIB BME) and percentage ([(FH BME − STBIB BME)/ STBIB BME] × 100) differences. These were also derived as both directional (i.e., maintaining positive and negative signs) and absolute values, which can be taken to represent systematic directional and random differences, respectively. Because of the way in which they are calculated, a positive directional difference indicates a relatively larger FH BME than STBIB BME, and vice versa. Average differences were calculated over the entire pooled sample, for different size ranges within the sample (see Results), and for each of the 17 narrower regional groups. The significance of directional differences among techniques was assessed using paired *t*-tests. For the entire pooled data set, 95% confidence intervals were also calculated for directional differences.

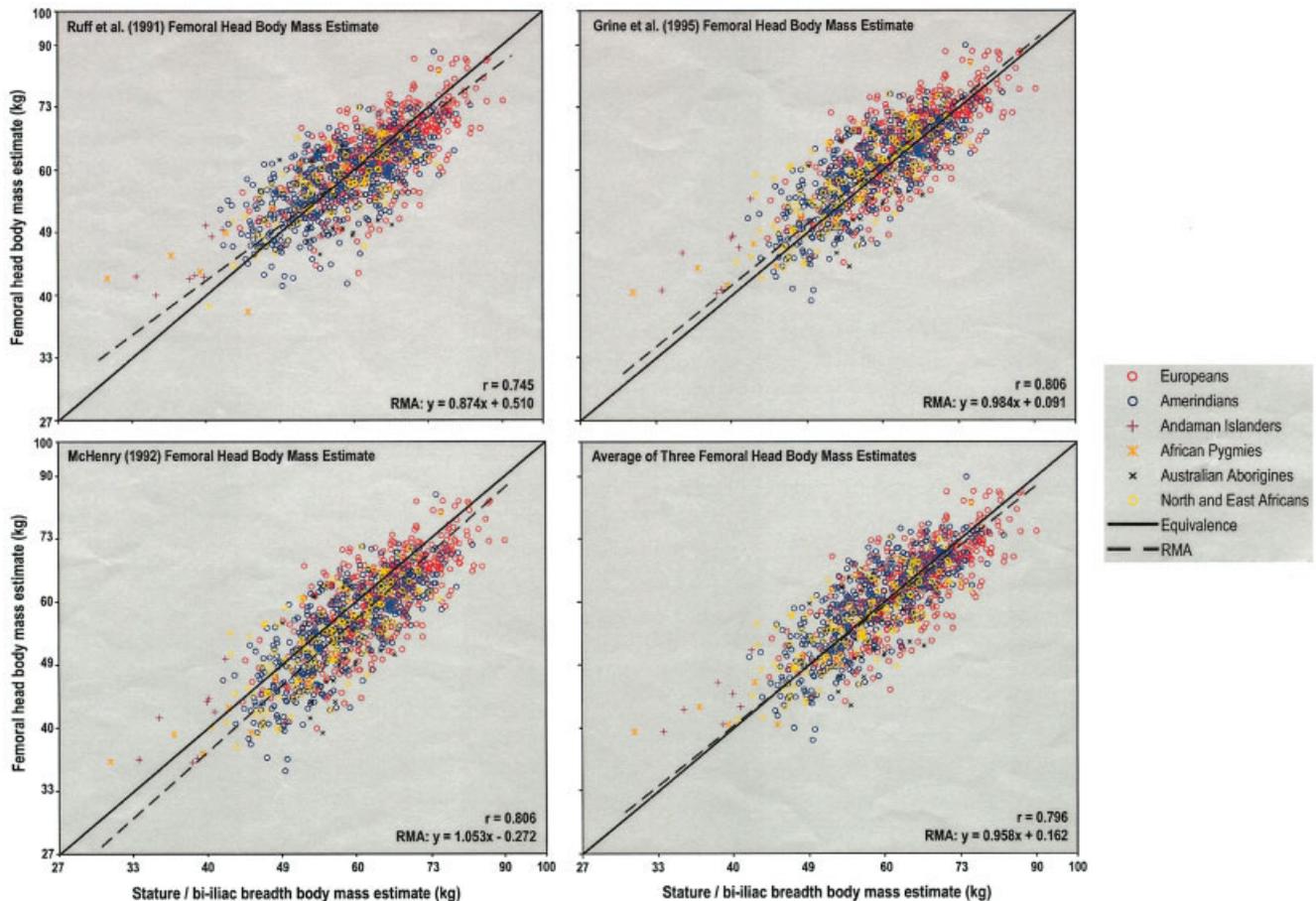[1]Note that the FH BME formula of McHenry (1992) is based on the raw data presented in his Table 1.

Fig. 1. Scatterplots of three femoral head body mass estimate techniques, individually and combined, against stature/bi-iliac breadth estimated body mass technique of Ruff et al. (1997). Solid lines demonstrate line of equivalence between technique estimations. Dashed lines are reduced major axis regressions of each femoral head technique on stature/bi-iliac breadth technique. Symbols are explained at right. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Because of the highly skewed distributions of absolute differences, for these differences, medians and lower 95% ranges are reported rather than confidence intervals.

All calculations were performed using Microsoft Excel XP for Windows and SPSS 10.0 for Windows. Graphs were produced in Microsoft Excel XP for Windows and Adobe Photoshop 6.0 for Windows. A color version of Figure 1 is available in the online version of this paper, which may be accessed at http://www.interscience.wiley.com or from the authors.

RESULTS

Data scatters for each of the three FH BME techniques plotted against the STBIB BME technique are shown in Figure 1. In addition, the mean of the three FH BMEs for each individual is plotted against the STBIB BME. Individuals from the six broad regional groups (Table 1) are plotted with different symbols, although RMA regressions were calculated through the total pooled sample. Lines indicating equivalence between techniques are also included in the plots for comparison.

Correlations between the FH and STBIB techniques range between 0.74–0.81. Overall, relative to the STBIB technique, the FH technique of McHenry (1992) tends to underestimate body mass, and the FH technique of Grine et al. (1995) slightly overestimates body mass. Comparison of RMA slopes to isometry (1.0), using a test described by Hofman (1988), indicates some size-dependence of technique differences for the FH comparisons of McHenry (1992) and Ruff et al. (1991), but not the FH comparison of Grine et al. (1995) or the comparison using the mean of the three FH estimates. The RMA slope is less than 1.0 for the comparison of Ruff et al. (1991), with the FH technique tending to overestimate body mass in smaller individuals and underestimate body mass in larger individuals compared to STBIB estimates. The RMA slope is greater than 1.0 for the comparison of McHenry (1992), with a greater difference between techniques in the smaller size range. However, for the very smallest individuals in the sample (African Pygmies and Andaman Islanders) the FH technique of McHenry (1992) appears to give estimates closer to those using the STBIB technique (i.e., points are closer to the line of

TABLE 3. Mean raw differences and median percent differences in femoral head body mass estimations vs. stature/bi-iliac breadth body mass estimates, all samples and sexes combined

Femoral head BME technique	Directional difference[1]		Absolute difference[2]	
	Raw (kg)	%	Raw (kg)	%
Ruff et al., 1991	−0.37* (−12.1, 11.4)	−0.15 (−19.9, 19.6)	3.98 (0.0, 11.9)	6.53 (0.0, 19.7)
McHenry, 1992	−3.14* (−13.9, 7.6)	−4.84 (−22.7, 13.0)	4.56 (0.0, 12.2)	7.41 (0.0, 19.3)
Grine et al., 1995	1.53* (−9.3, 12.4)	2.97 (−15.7, 21.6)	3.70 (0.0, 11.8)	6.02 (0.0, 14.6)
Average of 3 FH techniques	−0.66* (−11.5, 10.2)	−0.57 (−18.8, 17.7)	3.71 (0.0, 10.9)	6.08 (0.0, 18.5)

[1] Raw difference = FH BME − STBIB BME; % difference = [(FH BME − STBIB BME)/STBIB BME] × 100. Mean (95% CI).
[2] Absolute raw difference = |(FH BME − STBIB BME)|; absolute % difference = |[(FH BME − STBIB BME)/STBIB BME] × 100|; median (lower 95% range).
* Significant at $P < 0.05$ level by paired t-tests.

equivalence; Fig. 1). Otherwise, differences between techniques show no obvious systematic variation between the six broad regional groups (see below).

Mean directional and median absolute differences between the FH BME and STBIB BME techniques and the average of the three FH techniques are shown in Table 3, together with their 95% CIs (directional) or lower 95% ranges (absolute). Consistent with the data shown in Figure 1, the FH estimates of Ruff et al. (1991) and the mean of all three FH estimates show the least overall directional difference compared to STBIB estimates, while the FH estimates of McHenry (1992) are lower on average than the STBIB BMEs, and the FH estimates of Grine et al. (1995) are higher. Raw deviations for all FH techniques are significantly different from 0 ($P < 0.05$) using paired t-tests. However, average directional differences are relatively small (maximum about 3 kg, or 5%) in all comparisons. The 95% CIs for directional differences vary by technique, but generally fall between ±12 kg, or ±20%.

In contrast to directional differences, absolute (random) differences are similar for all three FH techniques (Table 3). This is again consistent with the data scatters in Figure 1. Median random differences average about 4 kg (6–7%) for each technique, with 95% of values falling below about 12 kg (20%).

Percent directional differences between FH and STBIB BMEs, classified by sex and broad and narrow groups (Table 1), are presented as box plots in Figure 2. As expected, given the earlier presented results (Table 3), the great majority of individuals fall within ±20% of 0, i.e., of equivalence between FH and STBIB techniques. Although there is variation among groups, with few exceptions, all group median differences fall within ±15% and interquartile ranges fall within ±20% of equivalence. The major exceptions involve the smallest-bodied samples (female African Pygmies and Andaman Islanders) for which FH estimates are much greater than STBIB estimates using the FH formulae of Ruff et al. (1991) and Grine et al. (1995) (but not the FH formula of McHenry, 1992). Male Andaman Islanders (but not the single male African Pygmy) also show a large difference between technique results using the FH estimates of Grine et al. (1995), with less difference using the other two FH techniques.

Two other groups have an interquartile range exceeding ±20% or a median exceeding 15%, both using the FH equation of McHenry (1992): female Australian Aborigines and female Utes. These, like almost all of the FH results of McHenry (1992) (except Pygmoids), are underestimates relative to the STBIB BMEs. Consistent with Figure 1, the majority of groups have higher FH body mass estimates using the formula of Grine et al. (1995), while (smaller) females are generally overestimated, and (larger) males are underestimated using the formulae of Ruff et al. (1991). Other than the Pygmy and Andaman Islander results, there are no systematic differences between the six broad geographic groups. Australian Aborigine females tend to be underestimated using the FH BME techniques, although this is one of the smaller samples (n = 9) in the study; the somewhat larger male Australian Aborigine sample (n = 15) shows no such consistent deviation. There is some patterning to the variation between more narrowly defined groups, e.g., compare the relative positions of groups 15–17 (Pueblo, Santa Cruz Islanders, and Utes) in each plot in Figure 2.

There is also some suggestion of a general similarity between males and females in the pattern of relative differences between narrow groups, at least among Amerindians (where "narrow group" corresponds more closely to true biological populations; see Table 1). The possible factors underlying these results are beyond the scope of the present paper. They suggest, however, that similar physiological/biomechanical mechanisms (e.g., femoral head size relative to body size, or systematic differences in linear body proportions affecting stature reconstruction) may operate across sex within populations (see Sjøvold, 1990).

Fig. 2. Box-plots of ranges of difference between femoral head body mass estimation techniques and stature/bi-iliac breadth estimated body mass technique of Ruff et al. (1997), by femoral head estimate technique, sex, and population. Whiskers are at quartiles from population medians (black horizontal bars in boxes). Broad groups are defined by pattern (see Table 1): left to right diagonals, Andaman Islanders; shaded, African Pygmies; dotted, Australian Aborigines; cross-hatched, East Africans; right to left diagonals, North Africans; solid, Europeans; no pattern, Amerindians. Narrow groups are enumerated as listed in Table 1.
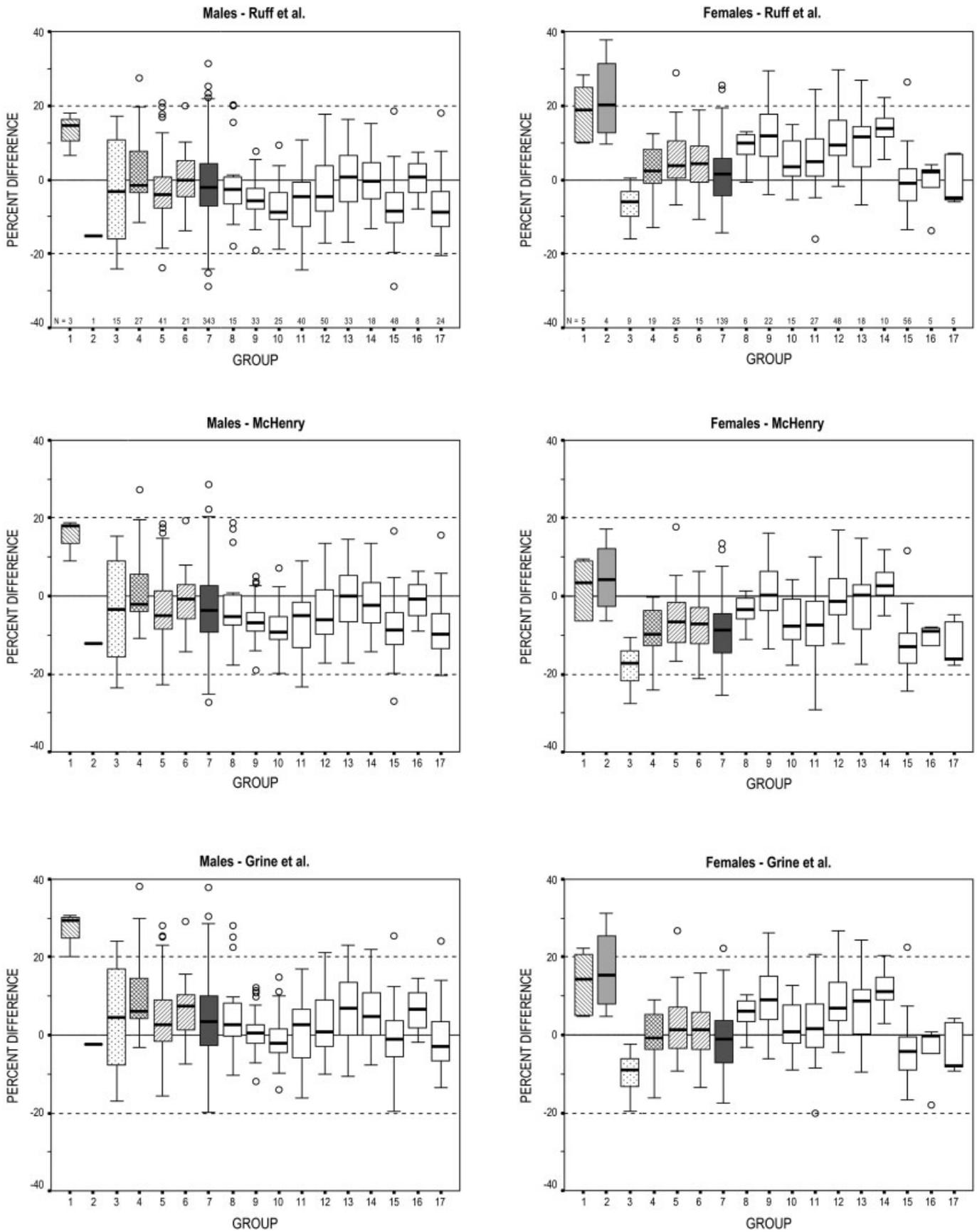
Fig. 2.

TABLE 4. Mean differences in femoral head body mass estimates vs. stature/bi-iliac breadth body mass estimates, by body size range

| Femoral head BME technique | Size range group | Directional difference[1] | |
		Raw (kg)	%
Ruff et al., 1991	Pygmoid females[2]	7.38 kg*	19.98%
	Smaller non-Pygmoid[3]	2.21 kg*	4.33%
	Larger non-Pygmoid[4]	−3.15 kg*	−4.49%
McHenry, 1992	Pygmoid females	1.03 kg	3.17%
	Smaller non-Pygmoid	−1.80 kg*	−3.27%
	Larger non-Pygmoid	−4.62 kg*	−6.67%
Grine et al., 1995	Pygmoid females	5.45 kg*	14.87%
	Smaller non-Pygmoid	2.79 kg*	5.27%
	Larger non-Pygmoid	0.12 kg	0.31%
Average of 3 FH techniques	Pygmoid females	4.08 kg*	10.67%
	Smaller non-Pygmoid	1.06 kg*	2.11%
	Larger non-Pygmoid	−2.55 kg*	−3.62%

[1] Mean raw difference = FH BME − STBIB BME. Mean % difference = (FH BME − STBIB BME)/STBIB BME × 100.
[2] Combination of African Pygmies (n = 4) and Andaman Islanders (n = 5). Females are used only due to small male sample sizes and extreme small size of pygmoid females when compared with other human populations.
[3] Defined as all non-pygmoid individuals under overall mean weight of 60.88 kg.
[4] Defined as all non-pygmoid individuals over overall mean weight of 60.88 kg.
* Significant at $P < 0.05$ level by paired *t*-tests.

Because there is evidence for body size-related variation in the correspondence between FH and STBIB BME technique results, mean directional differences between techniques were also calculated over different broad body size ranges. Three ranges were considered: female "Pygmoids" (African Pygmies and Andaman Islanders), who represent the smallest individuals in the study sample (31.0–42.7 kg); other individuals below the overall grand mean body mass (40.7–60.8 kg); and other individuals above the grand mean body mass (60.9–84.9 kg) (all based on STBIB BMEs). Results are shown in Table 4. As expected, the FH formula of McHenry (1992) provides much closer estimates for the very small Pygmoid females, and results using his formula are not significantly different from the STBIB estimates by paired *t*-test ($P < 0.05$). The FH formula of Grine et al. (1995) provides the closest estimates for larger non-Pygmoids (again not significantly different from the STBIB BMEs at $P < 0.05$), and all three FH techniques have similar magnitudes of differences for smaller non-Pygmoids. The average BMEs over all three FH techniques give the smallest directional differences for smaller non-Pygmoids, and the second smallest directional differences for the other two size ranges.

DISCUSSION

Overall, body mass estimates based on the femoral head and on reconstructed stature and bi-iliac breadth show strong correspondence over a large range of Holocene humans, with little mean directional bias. Similar results were previously obtained for a much smaller sample of Pleistocene *Homo* (Ruff et al., 1997). These new results are encouraging, since the two types of methods are based on such different skeletal dimensions and underlying assumptions, and support the continued use of these formulae for body mass estimation.

However, despite the general correspondence between techniques, there were also some differences in results. Some of these may be attributable to differences among the original reference samples used to generate the formulae. The stature/bi-iliac equations were derived from a worldwide sample representative of the full range of variation in body size and shape among modern humans (Ruff, 1994). Thus, they are considered relatively unbiased in terms of reference sample composition. However, each of the femoral head equations was based on a more limited sampling of modern humans, which may affect their applicability in different situations.

The formulae of Ruff et al. (1991) were developed from a sample of 80 living Baltimore blacks and whites who ranged in body mass from 42–135 kg, completely encompassing (except for Pygmy and Andaman females) the current sample size range. With a mean of 77 kg, most individuals in the sample of Ruff et al. (1991) fell toward the larger end of the present study distribution. However, as argued in the original publication, average body mass in this sample had probably increased substantially since the cessation of skeletal growth, largely due to development of excess adiposity (mean age of the sample was 52 years). Average body mass at age 18 in this sample was 65 kg, not far from the mean estimated body mass of 61 kg in the present sample. It is likely that the largely preindustrial populations included in the present sample more closely approximated the 18-year-olds in terms of body composition and body mass relative to femoral head breadth (in fact, an adjustment for increased adiposity in the reference sample is built into the formulae; for discussion, see Ruff et al., 1991). Thus, the distribution of "mechanically relevant" body masses in the sample of Ruff et al. (1991) may not have been too different from that found in the present study. This may explain why these formulae have the best over-

all performance in terms of mean directional bias (compared to STBIB BMEs) of the three FH BME techniques. However, while the FH BME formulae of Ruff et al. (1991) perform well in the middle of the present study-sample body mass distribution and when averaged over the whole distribution, they perform more poorly near both the upper (underestimates) and lower (overestimates) ends of this distribution.

The FH BME formula of McHenry (1992) was designed to estimate body mass in (generally) smaller early hominins, and was based on four sample means for modern humans that spanned a small to medium body size range (30.4–64.9 kg). It was the only one of the three FH BME techniques to include African Pygmies (and also the small-bodied Khoisan). Consistent with this, the FH BME formula derived from the data of McHenry (1992) produced the closest correspondence with STBIB BMEs for our smallest individuals, i.e., female African Pygmies and Andaman Islanders.[2] This suggests that for very small humans, the formula of of McHenry (1992) may provide the best FH BMEs.

In contrast, the FH BME formula of Grine et al. (1995) was based on 10 sex/sample means specifically chosen to represent large-bodied modern humans, for use in estimating the body mass of the very large Berg Aukas proximal femoral specimen. Body masses in their sample ranged from 54–84 kg (W.L. Jungers, personal communication), corresponding approximately to the upper half of the range of the current sample. Of all the FH BMEs, the method of Grine et al. (1995) shows the least systematic deviation from STBIB BMEs in the larger size range of the current sample, i.e., the size range for which the equations were originally developed. Thus, for estimation of body mass specifically in large-bodied humans, the formula of Grine et al. (1995) may be the most accurate of the FH BME techniques.

The increasing differences from STBIB estimates at the estimation extremes of the FH BME method of Ruff et al. (1991), and some of the results for the other two FH BME equations, may be attributable to another general issue in prediction: the choice of which regression model to employ. Ordinary least squares regression (or "inverse calibration," in the terminology used by Konigsberg et al., 1998), was used to generate all of the original body mass estimation equations included here. This method is most appropriate when the sample to which the formula is applied is close to the center of the distribution of the reference sample, but may produce large differences near the upper and lower ends of this distribution (Sjøvold, 1990; Konigsberg et al., 1998). Least squares prediction is particularly inappropriate when extrapolating beyond the ends of the reference sample distribution, when alternative methods should be applied (Aiello, 1992; Konigsberg et al., 1998; Hens et al., 2000). These problems are exacerbated when correlation coefficients are relatively low. Partly because the FH BME formulae of both McHenry (1992) and Grine et al. (1995) were based on only a few sample mean data points, correlation coefficients were relatively high ($r > 0.90$), while those for the sex-specific formulae of Ruff et al. (1991), based on 41 male and 39 female individuals, were much lower ($r = 0.41–0.50$) (Ruff et al., 1997). Thus, the systematic deviations between body mass predictions at both upper and lower body sizes (but not near the mean body size) for the formulae of Ruff et al. (1991) may be a result of using least squares regression to calculate the original equations. The poor performance of all FH BME equations, except of McHenry (1992), for predicting the body mass of very small humans may be a result of extrapolating beyond the size range of these reference samples.

Averaging BMEs over all three FH techniques has some advantages, especially when body sizes are not near the extremes of the modern human distribution. Pooled over the entire size range of our sample, average FH BMEs were about equal to the best single FH techniques in both directional and absolute differences, and performed relatively well in all size ranges, as assessed through comparisons with STBIB BMEs. At the extremes of the modern human distribution, the averaging of the three techniques yielded more difference from the STBIB BMEs than the methods of McHenry (1992) and Grine et al. (1995) for the smallest and largest individuals, respectively. When sex of a skeletal specimen is uncertain and os coxae are not available, mixed-sex formulae for all of the FH BME equations are available (Table 2). As noted previously (Ruff, 2000a), however, the average of male and female equation STBIB estimates should be used in such circumstances.

Except for issues related to overall body size, there do not appear to be any systematic differences between FH and STBIB BMEs attributable to general geographic location or differences in body proportions. For example, generally warm-climate samples (East Africans or Australian Aborigines), which have more linear body proportions (Ruff, 1994), show similar patterns of differences between techniques as higher-latitude samples. Thus, there is apparently no general tendency for ecogeographically influenced body proportions to affect the relationship between FH and body mass (as assessed by STBIB BMEs). This is reassuring, since it implies that FH BME formulae should be applicable across

---

[2]In part this may be due to the fact that McHenry (1992) estimated body mass for his smallest samples (Pygmies and Khoisan) from reconstructed stature, using the same set of formulae (Olivier, 1976) that we did (except for one of the Pygmies). However, he then calculated body weight from stature for these, using a power relationship derived by Jungers and Stern (1983) from living Pygmies, which is different from the STBIB method. Thus the two BME techniques, while both "morphometric," are not equivalent (the correlation between BMEs using the stature/body mass power relationship vs. the STBIB method in our 13 Pygmies and Andaman Islanders is only $r = 0.325$, $P > 0.20$), reducing the possibility of circular reasoning.

a wide range of body shapes. On a finer scale, the patterning of differences among and within more narrowly defined groups here does imply some subtle effect of body proportions on technique results, but these apparent population-specific effects are beyond the resolution (and interest) of most archaeological and paleontological studies.

While STBIB BMEs were used here as a standard for comparison with FH BMEs, there is certainly some error associated with these estimates as well, as a consequence of both estimation error in the original formulae (Ruff, 2000a) and errors in reconstruction of stature (and the soft-tissue contribution to "living" bi-iliac breadth; see Table 2, footnote 5). Stature reconstruction is particularly problematic where stature estimation formulae for closely related groups are not available. We used the crural index as a guide to the appropriateness of different stature estimation formulae when no other evidence was available (see Appendix), but this is obviously an imperfect means of assessing the reliability of such techniques, and in some cases may not be possible (if tibial as well as femoral lengths are not measurable). Thus, body masses estimated using the STBIB technique will be subject to more potential error when stature formulae cannot be securely matched with the target specimen/sample. Of course, more error is introduced when bi-iliac breadth must be estimated rather than directly measured. However, if stature can be estimated with some confidence, and bi-iliac breadth is available, we consider the STBIB technique to be the most reliable of all BME methods evaluated here, since it is based on the widest sampling of modern humans and takes the most direct approach, i.e., it does not rely on any assumed mechanical relationship. This latter issue can be particularly problematic in earlier hominins, where locomotor mode and joint loadings may have been different than in modern humans, making choice of the most appropriate reference sample difficult (Jungers, 1988; McHenry, 1992; Ruff, 1995; Ruff et al., 1999).

Further testing of these methods in samples of known body mass would be advantageous in defining their limitations and true accuracies. Multiple regression using several skeletal parameters (bone lengths, articular breadths, and bi-iliac breadth) and known body masses could potentially lead to improved body mass estimation. However, at present, such data are not available for large, morphologically diverse samples of known body mass. We suggest that the FH and STBIB methods provide the best currently available techniques for estimating body mass in skeletal samples.

CONCLUSIONS

The good correspondence between three femoral head body mass estimation methods (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995) and the stature/bi-iliac method (Ruff et al., 1997) in a large, diverse sample of Holocene humans supports the use of both types of techniques, although the choice of technique(s) to employ will vary depending on the particular specimen/sample. We make the following general recommendations:

1) When bi-iliac breadth can be measured and stature reconstructed with some confidence (i.e., when appropriate stature reference samples are available), the STBIB technique should provide the most generally reliable BMEs.
2) If bi-iliac breadth is not available and/or stature estimates are questionable, FH formulae give reasonable (comparable) results. Unless the specimen/sample is from the large or small extremes in size relative to modern humans, either using the formulae of Ruff et al. (1991) or averaging of estimates from the three FH BME techniques is the best approach. If the specimen/sample is in the very small (Pygmoid) body size range, the formula of McHenry (1992) can be used instead; if it/they are in the very large body size range, the formula of Grine et al. (1995) can be used.
3) Because they are based on different underlying rationales, comparing estimates from both the FH method and the STBIB method can provide a useful check on each approach. Averaging of results from both approaches is recommended when there is any reason to question one or the other, e.g., when stature estimates are uncertain.

LITERATURE CITED

Aiello LC. 1992. Allometry and the analysis of size and shape in human evolution. J Hum Evol 22:127–147.

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

Allbrook D. 1961. The estimation of stature in British and East African males. J Forensic Med 8:15–28.

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. Anthropological papers no. 83. New York: American Museum of Natural History.

Feldesman MR, Lundy JK. 1988. Stature estimates for some African Plio-Pleistocene fossil hominids. J Hum Evol 17:583–596.

Flower W. 1889. Description of two skeletons of Akkas, a Pygmy race from central Africa. J R Anthropol Inst 18:3–19.

Genoves S. 1967. Proportionality of the long bones and their relation to stature among Mesoamericans. Am J Phys Anthropol 26:67–78.

Grine FE, Jungers WL, Tobias PV, Pearson OM. 1995. Fossil *Homo* femur from Berg Aukas, northern Namibia. Am J Phys Anthropol 26:67–78.

Hartwig-Scherer S. 1994. Body weight prediction in fossil *Homo*. Cour Forsch Inst Senckenberg 171:267–279.

Hens SM, Konigsberg LW, Jungers WL. 2000. Estimating stature in fossil hominids: which regression model and reference sample to use? J Hum Evol 38:767–784.

Hofman MA. 1988. Allometric scaling in palaeontology: a critical survey. Hum Evol 3:177–188.

Holliday TW. 1997. Body proportions in Late Pleistocene Europe and modern human origins. J Hum Evol 32:423–447.

Holliday TW. 2002. Body size and postcranial robusticity of European Upper Paleolithic hominins. J Hum Evol 43:513–528.

Holliday TW, Ruff CB. 1997. Ecogeographic patterning and stature prediction in fossil hominids: comment on Feldesman and Fountain. Am J Phys Anthropol 103:137–140.

Jungers WL. 1988. New estimates of body size in australopithecines. In: Grine FE, editor. Evolutionary history of the "robust" australopithecines. New York: Aldine de Gruyter. p 115–125.

Jungers WL, Stern JT. 1983. Body proportions, skeletal allometry and locomotion in the Hadar hominids: a reply to Wolpoff. J Hum Evol 12:673–684.

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

Konigsberg LW, Hens SM, Jantz LM, Jungers WL. 1998. Stature estimation and calibration: Bayesian and maximum likelihood perspectives in physical anthropology. Yrbk Phys Anthropol 41:65–92.

Krogman WM. 1962. The calculation of stature from long bones: the human skeleton in forensic medicine. Springfield, IL: Charles C. Thomas. p 153–187.

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

Man EH. 1883. On the aboriginal inhabitants of the Andaman Islands. London: Trübner & Co.

Mantegazza P, Regalia E. 1886. Studio sopra una serie di crani di Fuegini. Arch Anthropol Etnol 12:463–477.

Mathers K, Hennenberg M. 1995. Were we ever that big? Gradual increase in hominid body size overtime. Homo 46:141–173.

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

Olivier G. 1976. The stature of autralopithecines. J Hum Evol 5:529–534.

Phenice TW. 1969. A newly developed visual method of sexing the os pubis. Am J Phys Anthropol 30:297–301.

Porter A. 1995. The body weight of AL 288-1 ("Lucy"): a new approach using estimates of skeletal length and the body mass index. Int J Osteoarcheol 5:203–212.

Porter A. 2002. Estimation of body size and physique from hominin skeletal remains. Homo 53:17–38.

Robins G, Shute CCD. 1986. Predynastic Egyptian stature and physical proportions. Hum Evol 1:313–324.

Ruff CB. 1988. Hindlimb articular surface allometry in Hominoidea and *Macaca*, with comparisons to diaphyseal scaling. J Hum Evol 17:687–714.

Ruff CB. 1994. Morphological adaptation to climate in modern and fossil hominids. Yrbk Phys Anthropol 37:65–107.

Ruff CB. 1995. Biomechanics of the hip and birth in early *Homo*. Am J Phys Anthropol 98:527–574.

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

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

Ruff C. 2002. Variation in human body size and shape. Annu Rev Anthropol 31:211–232.

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

Ruff CB, Walker A, Teaford MF. 1989. Body-mass, sexual dimorphism and femoral proportions of *Proconsul* from Rusinga and Mfangano Islands, Kenya. J Hum Evol 18:515–536.

Ruff CB, Scott WW, Liu A Y-C. 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 CB, McHenry HM, Thackeray JF. 1999. The "robust" australopithecine hip: cross-sectional morphology of the SK 82 and 97 proximal femora. Am J Phys Anthropol 109:509–521.

Ruff CB, Trinkaus E, Holliday TW. 2002. Body proportions and size of Lagar Velho 1. In: Zilhao J, Trinkaus E, editors. Portrait of the artist as a child. The Gravetian human skeleton from the Abrigo do Lagar Velho and its archaeological context. Lisbon: Instituto Portugues de Arqueologia. p 365–391.

Sciulli PW, Giesen MJ. 1993. Brief communication: an update on stature estimation in prehistoric Native Americans of Ohio. Am J Phys Anthropol 92:395–399.

Sjøvold T. 1990. Estimation of stature from long bones utilizing the line of organic correlation. Hum Evol 5:431–447.

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

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

Trotter M, Gleser GC. 1952. Estimation of stature from long bones of American whites and Negroes. Am J Phys Anthropol 10:463–514.

Trotter M, Gleser CG. 1958. A re-evaluation of estimation of stature based on measurements of stature taken during life and of long bones after death. Am J Phys Anthropol 16:79–124.

van der Broek A. 1940. Das Skelett einer weiblichen Efe-Pygmae. Z Morphol Anthropol 38:122–174.

APPENDIX

Notes on stature estimation

Stature formulae, as listed in Table 2, were applied as explained in Methods. Generally, groups from this study were either matched with reference samples or groups with similar body proportions, as assessed by similarity in crural indices (tibial length/femoral length). Here we review the specific methods chosen for stature estimations for each narrowly defined group (Table 1).

The Baka Pygmy stature formula of Olivier (1976) was applied to both the African Pygmies and Andaman Islanders. The similarity of overall body proportions of the Andamans to African Pygmies is not known, but similar crural indices (see Methods) and Pygmoid size argue for the use of the same equation. Additionally, the mean estimated stature/bi-iliac breadth body mass estimates of Andaman Islanders are nearly identical to population means calculated from anthropometrics recorded by Man (1883); though somewhat tautological, this further supports the use of the formula of Olivier (1976) for the Andaman Islands' Pygmoids.

The Nilohamite equation of Allbrook (1961) was used to estimate statures for East Africans and Native Australians. East Africans in the present study sample are most closely matched to the populations observed by Allbrook (1961) both geographically and in terms of body proportion (see Ruff and Walker, 1993). Australian Aborigines' crural indices overlap more with the East Africans of Allbrook (1961) than with other potential reference populations (e.g., South Africans observed by Feldesman and Lundy, 1988).

The formulae of Trotter and Gleser (1952) were applied for Europeans, Alaskan Inuit, Aleut, and

North Africans. Europeans in this study were universally estimated with the "white" stature formulae of Trotter and Gleser (1952). The equations of Trotter and Gleser (1952) were chosen over those of Trotter and Gleser (1958), due to the necessity of sex-specific formulae. These "white" equations were also used in the calculation of Alaskan Inuit and Aleutian Islanders, as the very low crural indices of the Amerindians overlap with the crural indices of the sample population used by Trotter and Gleser (1952. Following Holliday (2002; see also Ruff et al., 1997), statures of the Egyptian and Nubian populations were estimated by calculating statures with both the "white" and "negro" formulae of Trotter and Gleser (1952), and then averaging these estimated statures. Predynastic Egyptian stature equations developed by Robins and Shute (1986) were tested against the averaging method of Holliday (2002), and showed good correspondence (an average of 1.1 cm difference between techniques). We preferred the technique of Holliday (2002) to better correspond with the sample included herein, as the Egyptians used in this study are Dynastic, and Nubians were not incorporated into the calculations Robins and Shute (1986).

Amerindian statures, other than the previously noted Inuit and Aleut, were computed according to Genoves (1967) or Sciulli and Giesen (1993). The equation of Genoves (1967) for Mesoamericans was applied to all southwestern North American groups (Pueblo, Santa Cruz Island, and Ute populations), as well as Terra del Fuego's Yamana. Though it seems unlikely to use a Mesoamerican stature formula for Fuegians, the Yamana do have similar crural indices to the sample used by Genoves (1967), and mean stature estimates, by sex, using Genoves (1967) are not exceptionally different (less than 3 cm) from living heights of the Yamana obtained from the literature (Mantegazza and Regalia, 1886). The stature equations of Sciulli and Giesen (1993) were developed for Eastern Woodlands Amerindians from Ohio, and match the crural indices and general geographic range of the remaining populations: Libben, Indian Knoll, Arikara, and Illinois Hopewell.

ERRATUM: B.M. Auerbach and C.B. Ruff (2004) Human body mass estimation: a comparison of "morphometric" and "mechanical" methods. Am. J. Phys. Anthropol. *125*;331–342.

Figure 1 on page 335 was published incorrectly and should be replaced with:
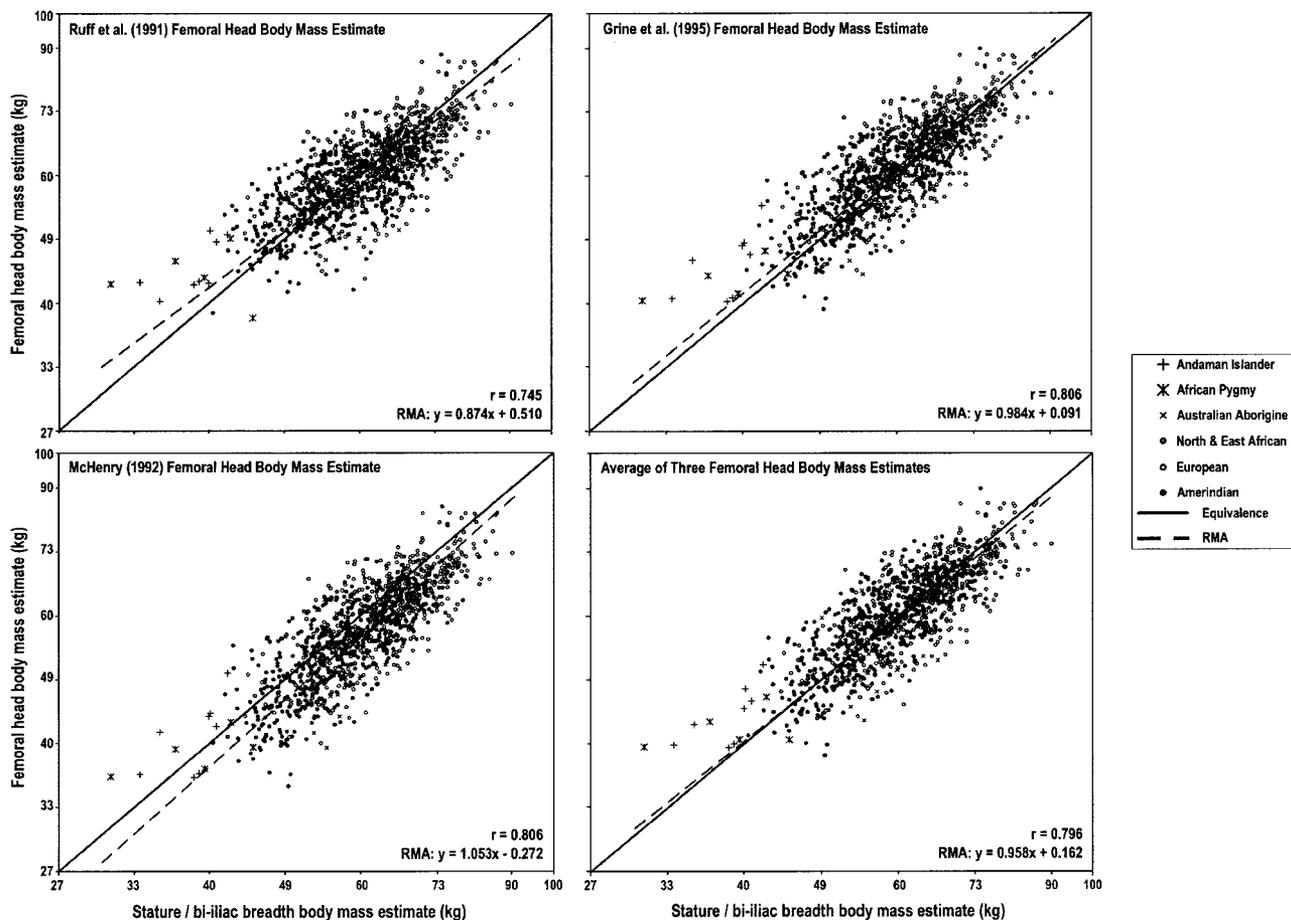

Figure 1.

A high-quality color version of Figure 1 is also now available from the authors' website: www.hopkinsmedicine.org/fae/.

Additionally, the Grine et al. (1995) reference was incorrectly cited in the Literature Cited section, and should be:

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