Estimating Body Mass In Biological Anthropology: An Evaluation Using Three-Dimensional Computed Tomography

by

Marina C. Elliott

M.A., Simon Fraser University, 2008
B.A., University of British Columbia, 2005

Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in the

Department of Archaeology

Faculty of the Environment

SIMON FRASER UNIVERSITY

Spring 2015

Approval

Name:	**Marina C. Elliott**
Degree:	**Doctor of Philosophy**
Title:	***Estimating Body Mass In Biological Anthropology: An Evaluation Using Three-Dimensional Computed Tomography***
Examining Committee:	**Chair:** Dana Lepofsky Professor

Mark Collard
Senior Supervisor
Professor

Helen Kurki
Supervisor
Associate Professor
Department of Archaeology
University of Victoria

Darlene A. Weston
Supervisor
Assistant Professor
Department of Archaeology
University of British Columbia

Hugo Cardoso
Internal Examiner
Assistant Professor
Department of Archaeology
Simon Fraser University

William Jungers
External Examiner
Professor
Department of Anatomical Sciences
Stony Brook University

Date Defended/Approved: January 9, 2015

Partial Copyright Licence

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010

Abstract

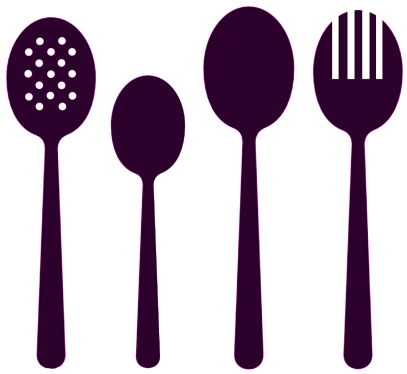
Estimates of body mass are essential to biological anthropology research. The primary source for such estimates is skeletal morphology, and several predictive equations have been developed for cranial and postcranial material. These equations are widely used, but a number of factors suggest that they may not be as reliable as they are generally assumed to be. In particular, reference samples are often small and analyses frequently employ indirect measurements, specimens without accompanying body mass values, or mean data. In addition, tests of the equations have rarely involved external validation with samples of known mass.

This project addressed these issues through three studies, using a large sample of modern humans for which both body masses and skeletal measurements were available. The sample consisted of Swiss forensic cases whose skeletal measurements were reconstructed from whole-body computed tomography scans. The first study compared the accuracy of three sets of commonly employed cranial equations. The second assessed published postcranial equations and compared the results to previous evaluations that had used less robust test samples. Several expectations regarding the performance of the equations were also tested. The third study employed the same sample to develop and test new regression equations for estimating mass from cranial and postcranial variables. The study was designed to compare the relative utility of the cranial and postcranial equations and to test the effect of variable choice, statistical method, and evaluation criteria on estimation competence.

Results suggest that body mass estimates should be used more cautiously than is usually the case. Overall, cranial equations did not estimate mass accurately. Several that have been deemed to be reliable in previous studies, did not perform well. Postcranial equations estimated mass more accurately, but not consistently. They also did not necessarily perform in accordance with statements in the literature. Deriving new equations using a known reference sample improved estimation competence compared to previous studies, but accuracy rates remained relatively low. Key assumptions about the best criteria to use for evaluating predictive competence were not supported. Further research may explain these discrepancies, but until then, estimates generated with currently published equations should be treated as “ballpark figures”.

Keywords: body mass; hominin; paleoanthropology; osteology; human variation, regression analyses

To The Spoons – you know who you are and what you mean to me.



Acknowledgements

Words cannot express how grateful I am to have the family, friends, colleagues and mentors that I have had during this process. First, I am indebted to my supervisor, Mark Collard, for his expertise, dedication and support - academic, financial and otherwise. His tireless commitment to scientific rigor encouraged me to strive for excellence in my research and made me a better scholar. Thanks also to my committee members, Helen Kurki and Darlene Weston, whose contributions and assistance were invaluable.

I thank Michael Thali and the staff at the Institute for Forensic Medicine, for providing access to the data I used for this research, as well as their assistance, patience and kindness during my stay in Zurich.

I am also indebted to the entire Human Evolutionary Studies Program team. They – and the Thursday Collective in particular – generously shared their knowledge, ideas and time and kept me excited about asking interesting questions and making a contribution to science. They were also the worst enablers and best therapy I could ever hope for. Another special thank you goes to Allison Spurrell and the Les Amis gang. Without their much needed humour, perspective and tea treats, my sanity would not have survived, if indeed it has.

Finally, I must thank my family and friends. I cannot possibly name them all, or list the innumerable ways in which they have helped me. But it has truly taken a village to raise this child and I could not have done it without them. One name I cannot omit, however, is my husband, Robin Elliott. Again, there are no words for what he has given me, and what he is to me. I only hope I can return the favours one day.

Science is a process – one that involves numerous starts, stops, triumphs and failures. But it is also a team effort, and without the collaboration, co-operation and assistance of these people (and many others) this research would not have happened at all. I am honoured - and humbled – every time I think about all that has gone into this endeavour. I have truly stood on the shoulders of giants, and at the same time, feel as if

I have stood beside many of them too – it has been amazing and I can only imagine what surprises the future holds.

Thank you,

Marina C. Elliott

This research and dissertation were generously funded by the Social Sciences and Humanities Research Council (#767-2009-1887 3), the Human Evolutionary Studies Program and Simon Fraser University.

Table of Contents

List of Tables

List of Figures

Chapter 1.

Introduction

1.1. Aims and Objectives

Body mass estimation is one of the most important tasks undertaken by biological anthropologists. As body mass correlates closely with, and indeed determines, many physiological, behavioural, and ecological traits (Calder, 1984; Damuth and McFadden, 1990), estimates from skeletal material are essential to investigations of past species and groups. In palaeoanthropology, estimates are used to infer traits as diverse as hair density, gestation length, inter-birth interval, positional repertoire, home range size, and dietary preference (Smith, 1996). Body mass estimates are also used in comparative studies and provide a baseline for assessing evolutionary changes and adaptive shifts in features like brain size, tooth size, limb proportion and neonatal size (e.g. Frayer, 1984; Delson et al., 2000; DeSilva, 2011; Frayer, 1984). In more recent humans, body mass estimates are essential for reconstructing demographic characteristics and assessing health in individuals and groups (Steckel and Rose, 2002; Cohen and Crane-Kramer, 2007). Body mass estimates also play a vital role in understanding patterns of growth and development (Humphrey, 1998; Walker et al., 2006) and interpreting physical responses to environmental influences (Ruff, 2002; Kurki et al., 2010; Cross and Collard, 2011; Kurki et al., 2010; Collard and Cross, in press). Lastly, body mass estimates are key to understanding biomechanical influences on the body and determining how genetic and epigenetic factors, lifestyle, and activity patterns shape skeletal morphology (Ruff, 2000; Lieberman et al., 2001; Ruff, 2000; Ruff et al., 2006).

Body mass estimates are also beginning to play an important role in forensic anthropology. Given the current rise in obesity around the world, the potential for determining differences in body mass from skeletal material in modern groups is of

interest to forensic anthropologists (Rainwater et al., 2007; Agostini and Ross, 2011; Daneshvari, 2011; Moore and Schaefer, 2011; Lorkiewicz-Muszyńska et al., 2013). This research is important not only because weight is an indicator of general health, but also because it may be possible to use body mass as an individualizing feature in mass disaster or identification work (Byard, 2012). In addition, because body size can affect decomposition, deposition and diagenic processes, body mass estimates could be used to help interpret taphonomic changes in a variety of contexts (Suskewicz, 2004).

In light of their obvious importance, it is imperative that the methods used to estimate body mass in biological anthropology are as robust as possible. In particular, the objectives and limitations of the methods must be understood and made explicit. Where predictive equations have been developed, and are used regularly, it must be demonstrated that they provide reasonable and accurate predictions for the specimen(s) in question. In addition, expectations of performance and levels of precision must be defined if any level of confidence is to be placed in the resulting mass estimate. Where such equations are shown to be problematic, new equations or approaches must be considered and interpretations must be appropriately tempered. Thus, it is critical to ensure that the methods are evaluated as thoroughly as possible. This was the impetus for the present research.

This project used whole-body computed tomography scans of known individuals to reconstruct three-dimensional models of specific skeletal regions commonly used to estimate body mass. Measurements from the reconstructions were then entered into previously published cranial and postcranial regression equations to test the accuracy of their estimates against the sample individuals' known masses. A number of expectations regarding the performance of the regression equations were also tested. The measurements were used to generate new regression equations for both cranial and postcranial variables. The accuracy of the new equations was then tested using a sub-sample of the reference group. The results of the cranial and postcranial analyses were compared to each other and to those of existing equations. Several factors thought to affect estimation accuracy were also assessed. In taking this approach, the present research sought to improve the way biological anthropologists estimate human body mass from skeletal material.

The remainder of this chapter provides a background on body mass estimation in biological anthropology and the current project's contribution to it. It begins by reviewing the history of body mass estimation and the different skeletal regions that have been used in past analyses. It then reviews in detail, the cranial and postcranial studies whose regression equations are now used most widely. The limitations associated with the different approaches and the equations derived from them, are also outlined. One of the more significant of these problems relates to the inadequacy of many reference samples. Consequently, the subsequent section discusses how the present research employed Computed Tomography to address this issue and the advantages of this approach. It concludes with a brief outline of the dissertation and the three papers that comprise it.

1.2. History of body mass estimation in Biological Anthropology

Attempts to estimate the size of past individuals and species have a long history in biological anthropology (Duncan, 1883; Pearson, 1899). While a variety of parameters (e.g. stature, trunk volume, bone weight, etc.) have been used to represent "size", body mass, as a dimension exclusive of force (Smith and Jungers, 1997), has been the primary focus for most studies.

Many early estimates of body mass for fossil primates, including hominins, focused on dental material (e.g. Henderson and Corruccini, 1976; Gingerich, 1977; Wood, 1979; Blumenberg, 1984; Conroy, 1987; Fleagle and Kay, 1995). This was an obvious choice as teeth preserve well, do not grow continuously in most mammals, and show strong heritability (Gingerich, 1977). Teeth are also well represented across time, space and taxonomic groups, making them useful foci for evolutionary studies (Wood, 1979). However, using teeth is problematic in fossil hominin studies as behavioural specializations (e.g. dietary preferences in robust australopithecines) can result in different body mass-tooth size relationships and thus, have the potential to bias the resulting mass estimate (Steudel, 1980; McHenry, 1988). In addition, the relationship between dental dimensions and body size in modern humans has proven to be relatively weak (Henderson and Corruccini, 1976). While this may be due to the increased reliance

on tools and food processing (Frayer, 1984; Kappelman, 1996), it further reduced the utility of dental material for body mass estimates in early *Homo* species.

In light of these challenges, other areas of the cranium have been investigated for their utility in estimating body mass (Steudel, 1980; Martin, 1981; Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007). This approach takes explicit advantage of the fact that cranial material is relatively plentiful in the fossil record and forms the basis for identifying the taxonomic affinities which are necessary for comparative studies (Aiello and Wood, 1994). Several sets of predictive equations have been developed for use with fossil hominins (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007), all of which rely on allometric relationships across primates and employ interspecific regression analyses.

Three studies that have become the most widely cited sources for estimating body mass from cranial material are worth describing in more detail. The first is Aiello and Wood's (1994) study "Cranial variables as predictors of hominine body mass". Using mean body masses for non-human primates from the literature (Harvey et al., 1987), and human masses from autopsy records, Aiello and Wood collected a total of 15 cranial variables (linear measurements and area calculations) from a sample of 250 primates, including 24 modern humans from the Raymond Dart Collection. Analyses were carried out on log-transformed mean data for two samples – a broad "Simian" group and a narrower "Hominoid" group, using least square (LS) and reduced major axis (RMA) regression. The cranial dimensions were also compared to 14 postcranial variables commonly used in body mass estimation. The results suggested that several cranial variables correlated well with body mass. In particular, orbital area, orbital height and biporionic breadth were argued to be the most reliable estimators of mass across hominoids, including humans. Body masses of several fossil specimens were then estimated using these variables, with the conclusion that "orbital height should be the preferred cranial predictor" for fossil hominins (Aiello and Wood, 1994:424).

The second study is Kappelman's (1996) "The evolution of body mass and relative brain size in fossil hominids". Here, orbital area was measured on 343 extant primates, including 32 humans from the Hamann-Todd collection. Again, mean body masses for the non-human primates were gleaned from the literature while the human

body masses were taken from autopsy records. Orbital area was measured from photographs imported into a Computer Assisted Design (CAD) program. As with Aiello and Wood's (1994) study, two samples, an "all primate" group and a "hominoids-only" group, were analyzed. All data were log transformed and analyses carried out on mean values using least square (LS) regression. Reduced major axis (RMA) regressions were conducted only for comparative purposes. The results were also compared to those derived from postcranial material, specifically femoral head measurements. Kappelman (1996) also found a strong correlation between orbital area and body mass (r=0.987) and considered his CAD-derived measurement method for orbital area to be an improvement on that of Aiello and Wood (1994). Like Aiello and Wood (1994), Kappelman applied his equations to a number of fossil specimens.

The third study to use cranial variables to approach the question of body mass estimation is "The use of cranial variables for the estimation of body mass in fossil hominins" by Spocter and Manger (2007). This study employed many of the same variables as Aiello and Wood (1994) and Kappelman (1996), but used slightly different calculations for some variables (e.g. orbital area). The sample consisted of 259 primates, including 180 modern humans from the Raymond Dart Collection, the same collection used by Aiello and Wood (1994). Again, body masses for the human sample were taken from autopsy records, while the non-human primate masses were taken from the literature. Analyses were conducted on mean data for two samples: all primates and hominoids. However, unlike the previous two studies, only average female body masses were used. Data were log transformed and analyzed using least square and reduced major axis methods. This study also differed from the previous two in explicitly taking phylogeny into account via the independent contrasts method (Felsenstein, 1985). Spocter and Manger (2007) found that orbital height, orbital area and facial breadth were the best predictors of mass. Subsequently, they used these features to estimate the mass of a range of fossil specimens.

Despite these three studies arguing in their favour, cranial variables have been considered to be less appropriate for body mass estimation than postcranial features, as the skull does not transmit the body's weight and therefore bears no functional relationship to mass (e.g. Hylander, 1985). On this basis, postcranial elements have

been the primary focus for most body mass estimation studies (e.g. McHenry, 1976; 1988; Rightmire, 1986; Ruff, 1987).

As the largest weight-bearing bone in the body, the femur has dominated postcranial body mass estimation methods. Initially, femoral shaft breadths were the primary target (McHenry, 1976; Oleksiak, 1986; Hartwig-Scherer, 1993). However, these efforts were criticized on the grounds that diaphyseal breadths can also be affected by changes in mechanical loading, environmental stress and activity (Ruff et al., 1993; Trinkaus et al., 1994; Lieberman et al., 2001). In addition, the relative size of diaphyseal breadths appears to differ between fossil and modern humans (Ruff et al., 1997), making comparisons between the two more difficult. As a result, later research has tended to focus on femoral articular surfaces.

Femoral head breadth (FHB) is now the most frequently used variable for estimating body mass and has become synonymous with the "mechanical approach" (Ruff, 2002). In addition to being considered less sensitive to external influences, this variable has been argued to be advantageous because femoral heads are relatively well represented in the fossil record and easily measured (Ruff et al., 1991). Three sets of regression equations derived from femoral head breadth have become standard in biological anthropology and are argued to be applicable to fossil and modern samples of varying body sizes (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995). An additional set of FHB equations, argued to be more broadly applicable to Holocene humans, has also recently been published (Ruff et al., 2012).

Given their broad use and general acceptance, the four current "mechanical" studies also warrant further discussion (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995; Ruff et al., 2012). In the first study, Ruff et al. (1991) tested the hypothesis that articular and diaphyseal morphologies respond differently to mechanical loads by measuring femoral head, neck, mediolateral subperiosteal, and cortical breadth dimensions in a sample of 80 living individuals. The sample was evenly split between males and females, almost two-thirds of which were "white". Measurements were taken from clinical radiographs while current weight and weight-at-maturity (~18 yrs) were obtained through patient recall. From these data, Ruff et al. (1991) concluded that, overall, femoral head dimensions correlated better with weight at maturity than current

weight, while shaft dimensions showed the opposite tendency. Sub-samples showed more inconsistency, with some groups (e.g. “white females”) showing low correlations for all measurements. Despite these inconsistencies, Ruff et al. (1991) concluded that proximal femoral dimensions were the most useful for predicting individual body mass in modern humans and that earlier human samples could be reasonably estimated by applying a “downward adjustment of about 10%” (pg. 411) to account for increased adiposity in the modern reference sample. Ultimately, Ruff et al. (1991) provided three sets of regression equations for males, females and combined-sex samples for use with femoral head breadth. These equations have been used on both fossil and modern samples and continue to be used regularly to estimate body mass in biological anthropology (e.g. Auerbach and Ruff, 2004; Kurki et al., 2010; Pomeroy and Stock, 2012).

McHenry (1992) carried out the second “mechanical” study. Because their proportions differ, McHenry suggested that using relatively large-bodied modern human samples to estimate body mass in small-bodied fossil hominins was problematic. To address this, he measured 13 postcranial variables, including femoral head breadth, on a broad sample of extant primates, including humans. The human sample included both Europeans and two small-bodied African groups (Khoisan and “Pygmies”). For the most part, individual body masses were taken from field notes (wild-caught non-human primates) or from autopsy records (humans). Body masses for the two small-bodied human samples (n=8) however, were estimated from long bone lengths. Analyses were conducted on log transformed mean data, using least square (LS), major axis (MA) and reduced major axis (RMA) regression, with LS being considered the most appropriate. McHenry (1992) found that all the postcranial variables correlated highly with body mass, both across hominoids and within Homo sapiens. Within the human sample, the femoral shaft measurements correlated slightly better with weight than did the femoral head. Despite this, McHenry also concluded that hind limb joint sizes were better predictors of body mass in fossil hominins. Subsequent to this study, Ruff et al. (1997) used McHenry’s modern human data to derive a combined-sex regression equation for femoral head breadth specifically designed to be applicable to small-bodied humans and fossil hominins.

The third study to use femoral head breadth to estimate mass was carried out by Grine and colleagues (1995). Using data from Jungers (1990), the reference sample consisted of 10 sex-specific means for several large-bodied modern humans, including Europeans, African Americans and Native Americans. Least squares regression was again used and the data were not log transformed. Although the resulting combined-sex equation was designed to determine the size of a specific specimen (the Berg Aukas femur), it has since been recommended as a useful equation for estimating body mass in large-bodied hominins more generally (Ruff et al., 1997).

Ruff et al. (2012) provide the most recent set of mechanically-based regression equations. This study employed a large (n=1145) reference sample of geographically diverse archaeological specimens to derive sex-specific regression equations for femoral head breadth. Individual skeletal measurements were taken, but body masses were estimated using another postcranial approach: the "morphometric method" (Ruff, 2002) that involves measures of stature and bi-iliac breadth (see below). Arguing that Model II statistical techniques are more appropriate over broad size ranges, this study employed reduced major axis regression (RMA) instead of the more common least squares approach. Ruff et al. (2012:9) concluded that these new regression equations would be "broadly applicable to European Holocene populations" and potentially more appropriate in circumstances where equations for specific populations or body types were not available. As this set of equations was only recently published, they have yet to be incorporated into many investigations (but see Trinkaus et al., 2014).

A second approach to estimating mass from postcranial material has been described as the "morphometric method" (Ruff, 2000a; 2002; Auerbach and Ruff, 2004). Here, body mass is estimated by reconstructing overall body shape from stature (most often taken from femoral length) and a measure of body breadth (usually bi-iliac breadth) (Ruff, 1991; 1994; 2002; Ruff et al., 1997; 2005). Initially, a reference sample of living individuals from 56 populations around the world was used (Ruff, 1991; 1994). Variables consisted of mean anthropometric measurements for bi-iliac breadth and stature, combined with body mass data gleaned from various other studies. Although the measurement methods sometimes differed between sources, and not every study recorded both variables, Ruff (1991) attempted to "match" stature and bi-iliac breadth for each sample by using the same number of individuals in each group. Raw (unlogged)

data were used to derive sex-specific least square regression equations for estimating body mass in humans, after converting skeletal bi-iliac breadth to living breadth (Ruff et al., 1997). These equations have been argued to be reliable because the approach does not make biomechanical assumptions about the relationship of skeletal morphology to mass and employs a geographically diverse sample (Ruff 2000b; Ruff et al., 1997). The range of the sample was also used to argue that the equations are applicable even at the extremes of human variation (e.g. lean athletes) (Ruff, 2000a).

A subsequent study by Ruff et al. (2005) expanded on the original reference sample by adding two relatively large-bodied, high-latitude groups (Finnish males and females). New sex-specific regression equations were then derived that have been argued to be "more broadly applicable, particularly to tall and wide-bodied males" (Ruff et al., 2005:390) than the previous equations. While these equations were designed specifically to increase the representation of high-latitude populations and be used with similar groups, in practice, they have largely replaced the earlier equations (Ruff et al., 2006; 2012; Lorkiewicz-Muszyńska et al., 2013).

1.3. Limitations of estimation approaches

Despite the general acceptance of the methods, and the widespread application of the regression equations just described (Trinkaus and Jelinek, 1997; Kordos and Begun, 2001; Holliday, 2002; Ruff et al., 2006; Rosenberg et al., 2006; Ruff et al., 2006; Nakatsukasa et al., 2007; Knusel et al., 2010; Melton et al., 2010; Ruff, 2010), a number of practical, methodological, and statistical limitations have been identified in relation to estimating body mass from skeletal material.

Practically, one of the most obvious challenges relates to the choice of variable to employ for accurate estimation – a decision that depends heavily on what skeletal material is available. Complete elements are rare in the fossil record, and are not always common in archaeological or modern contexts (Haglund and Sorg, 2002). Consequently, measurements are frequently estimated from fragmentary or reconstructed elements, a practice that compounds the error in the resulting estimate. This is a challenge for the cranial equations as the "best" variables relate to the eye orbits and basicranium (Aiello and Wood, 1994), features that are easily broken or distorted in deposition. It is also a

problem for the “morphometric” postcranial equations as they require measures of both bi-iliac breadth and stature (Ruff, 2002; Ruff et al., 2005). Stature in particular, is fraught with numerous challenges (SWGANTH, 2012) whether it is “anatomically” reconstructed from a whole skeleton (Fully, 1956; Raxter et al., 2006), or estimated from a single element (e.g. Trotter and Gleser, 1952; Feldesman and Fountain, 1996).

The choice of variable also depends on the nature of its relationship to body mass. In general, the assumption has been that body mass should be better predicted by features that are directly related to weight (Jungers, 1988; Ruff, 1991, 1994; McHenry, 1992). For most researchers, this refers specifically to load-bearing elements like the femur. However, Smith (2002) has argued that while a functional relationship may suggest potential candidates for predictive competence, it is the statistical properties that determine a variable’s utility. Indeed, it is on this basis that the morphometric methods have been argued to be effective, despite the lack of a biomechanical relationship between mass and overall body shape (Ruff et al., 1997, 2005, 2012).

The choice of reference sample is also a critical issue (Smith, 2002). All methods for estimating body mass in biological anthropology rely on comparisons with extant primates. However, studies vary in terms of how many species are used and whether or not humans are included in the samples. Choices include a broad range of species (e.g. all primates), a narrower taxonomic group (e.g. only hominoids), or a single species (e.g. modern humans). This decision is based largely on whether or not the feature in question shares the same relationship to mass in each group (Steudel, 1980; Hartwig-Scherer, 1993) and whether the target specimen is expected to fall within the range of variation provided in the reference groups. Where humans and non-human primates scale the same way – as has been argued for orbital area (Kappelman, 1996) and femoral shaft breadth (McHenry, 1992), broad taxonomic comparisons are expected to provide a “greater degree of confidence” (Kappelman, 1996:260) in an estimate. Where humans scale differently from other primates, as with femoral head diameter because of bipedality (Ruff et al., 1991), the use of more narrow reference sample is considered more appropriate. For some equations, this has led to the exclusive use of modern human samples (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995).

In both cases, it is expected that the target specimen is a member of one of the populations that was used to generate the regression equation (Smith, 2002; Wood, 2011), or at least corresponds closely to the reference model in some biologically meaningful way (Hartwig-Scherer and Martin, 1992). However, it is not always possible to know how representative a reference sample is or to what extent a fossil specimen scales the same way as the extant groups (Hartwig-Scherer, 1993; Niskanen and Junno, 2009). Even within modern humans, body proportions and robusticity vary and predictive equations derived from one sample may not apply to another (Ruff et al., 1997; Ruff et al., 2012). Indeed, these differences have been used to argue in favour of population-specific equations for modern samples (Ruff, 1994). However, this again assumes the target population can be identified and matched to the reference group – an issue that will likely always be problematic (Ruff et al., 2012).

Another consideration relates to the use of individual data or population means to derive the regression equations for estimating mass. Individual, associated data are preferred in order to most accurately capture the relationship between a skeletal feature and body mass (Ruff et al., 1991; Niskanen and Junno, 2009). However, individual body masses are rarely recorded for non-human primates (Kappelman, 1996) and many "documented" human skeletal collections contain excessive numbers of emaciated individuals or suspect body masses (Stuedel, 1980). In addition, well-documented collections are often so small that they do not provide sufficient material for testing existing methods or developing new ones (Wood and Collard, 1996). Consequently, many studies base their regressions on mean data, either for several modern human populations (e.g. Ruff, 1991; Ruff et al., 2005) or multiple primate species (e.g. Aiello and Wood, 1994; Kappelman, 1996). While this allows for larger, more diverse samples to be included in the regressions, it clearly reduces the predictive power of the equation. Importantly, the assumption has been that if interspecies mean data are used, the resulting mass is that of the taxon and the value is only ever used to examine broad interspecies relationships (Gingerich, 1977; Hartwig-Scherer and Martin, 1992). However, this is not the case and the equations have been used to predict the mass of individual specimens, often within the same species (Aiello and Wood, 1994; Dagosta and Terranova, 1992; Kappelman, 1996; Kordos and Begun, 2001; Spocter and Manger, 2007).

The choice of which regression method to employ is significant as well. To date, the most common approach to body mass estimation has been Least Squares (LS) regression. This method predicts a value Y (e.g. body mass) from X (e.g. femoral head breadth) using a best-fit line that minimizes the sum of the squared deviations of the estimated values from the known values (Hartwig-Scherer and Martin, 1992). It has been argued to be the best way to predict one variable from another (Konigsberg et al., 1998; Smith, 1996, 2002), especially when the goal is to minimize the estimation error of the dependent variable (in this case, body mass) (Aiello and Wood, 1994). LSR also has the advantage of having correction factors available to account for biases inherent in the technique (Smith, 1996). However, LSR has been criticized for assuming the independent variable (i.e. the skeletal measurement) is measured without error. This is certainly not the case with reconstructed fossil elements or when different methods are used for the same measurement (see discussion below in relation to bi-iliac breadth). LSR has also been argued to produce biased results when applied outside the range of the sample from which it was derived (Konigsberg et al., 1998; Ruff et al., 2012).

In light of the putative problems with LSR, the Reduced Major Axis (RMA) method has been suggested as more appropriate for body mass estimation (Ruff et al., 1991; Auerbach and Ruff, 2004). RMA techniques allow both dependent and independent variables to be sampled with error and generally accept more uncertainty in the variable estimates. RMA techniques also produce a best-fit line "unaffected by the magnitude of the correlation coefficient" (Aiello and Wood, 1994:412). As a result, RMA techniques have been argued to produce better results when extrapolating beyond the range of the original dataset and may be applicable over a broader size range. Despite this, LSR techniques continue to be used regularly in body mass estimation research (e.g. Pomeroy and Stock, 2012) and a clear consensus on the best approach has not been reached.

The final challenge relates to how the predictive competence of an equation is evaluated, once it has been derived. Ideally, the most rigorous means of establishing a predictive equation's accuracy is through the use of external validation using a large, independent sample of known mass individuals (Porter, 1999; Giancristofaro and Salmaso, 2003). However, as noted earlier, very few reference collections meet these criteria and their small size has made it impractical – or impossible - to split samples into

the groups necessary to generate, and also test, the equations (Kappelman, 1996). As a result, predictive equations continue to be generated, and tested, using individuals whose body masses are not known (e.g. Pomeroy and Stock, 2012) – a circumstance that calls into question the accuracy of the resulting mass estimates.

In sum, a number of problems suggest that the equations currently available for estimating body mass from skeletal material may not be as reliable as they are assumed to be. Consequently, a systematic reanalysis of both the cranial and postcranial equations, and an exploration of the assumptions that underlie their use, was clearly warranted.

1.4. The application of CT technology

As just discussed, one of the primary issues in relation to body mass estimation in biological anthropology relates to the quality of the reference material. Existing skeletal collections are often small, lack associated cranial and postcranial material, and may not provide matched biological information for each individual. In seeking a way to address this problem, it became apparent that Computed Tomography (CT) might offer a potential solution.

Computed Tomography (CT), was first invented in the early 1970s (Hounsfield, 1973). Improving on previous radiographic technologies, CT uses multiple x-rays taken around the axis of an object to generate a cross-sectional slice of its three-dimensional form. X-rays passing through different tissues are absorbed and scattered differently, allowing small variations to be distinguished. Images are then produced on a computer screen by representing these density differences, or attenuation coefficients, as matrices of picture elements (pixels) (Hounsfield, 1980). Whole structures can be visualized by moving through consecutive slices or in three-dimensions by “stacking” slices using volume-rendering software. Like conventional radiographs, dense tissues are seen as lighter while less dense tissues are seen as darker (Fig 1.1). Since its development, CT has become the ‘gold-standard’ in medical diagnostic imaging (Beckmann, 2006).

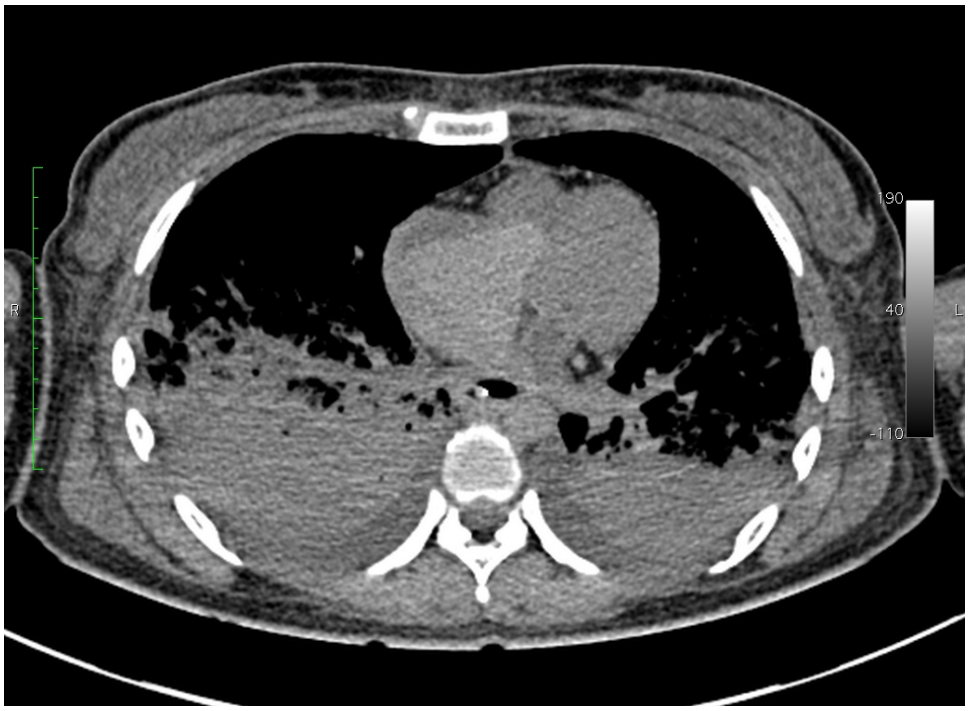


Figure 1.1. Example of a cross-sectional CT scan of the chest. Non-dense areas (air) are seen in black, while dense areas (bone) are seen in white.

CT has a number of advantages over other imaging techniques. First, CT is much better for visualizing and quantifying three-dimensional structures than two-dimensional media (like photographs or conventional radiographs). The spatial resolution and clarity of structures, especially bone, are better and there is very little distortion or magnification error (Adams et al., 2004). CT also avoids the problem of structural superimposition associated with conventional radiography (Reichs and Dorion, 1992). Lastly, as the technology has become more accessible and affordable, CT has become a viable tool for research outside the medical community.

Biological anthropologists were quick to recognize the value of CT for research (e.g. Jungers and Minns, 1979; Tate and Cann, 1982). Early studies used CT to visualize fossils still trapped in matrix (Wind, 1984), compare the internal morphology and structure of fossil and archaeological bones (Spoor et al., 1993) and to "virtually unwrap" ancient mummies (Lewin, 1988). More recently, CT data have been used to investigate disease processes in bone (Rühli et al., 2002), model changes in craniofacial shape over time (Zollikofer and Ponce De Leon, 2002) and to virtually reconstruct damaged fossils (Zollikofer et al., 2005). In forensic anthropology, CT has become the

standard in a number of jurisdictions for analyzing post-mortem remains without resorting to invasive procedures like autopsy, maceration or sectioning (Thali et al., 2003; Bassed et al., 2011). Combined with 3D printing technology, CT data are also being used to create casts for a number of purposes in biological anthropology (Tobias, 2001; Gunz et al., 2009; Kettner et al., 2011).

Advances in volume rendering and computer processing power now allow almost any structure to be converted into a three-dimensional virtual representation using CT data. These models can be constructed quickly, stored indefinitely, analyzed and re-analyzed without contact, and can even be shared by researchers across the world without the need for transporting fragile remains (Aghayev et al., 2008). More importantly for present purposes, modeled elements can be manipulated and measured accurately (Adams et al., 2004; Cavalcanti et al., 2004; Robinson et al., 2008). Given these advantages, CT data are being used to develop databases of material for reference and education purposes (e.g. Messmer et al., 2007). In particular, a number of forensic institutions have begun to archive whole-body CT scans taken as part of routine post-mortem examinations (e.g. Thali et al., 2003, Bassed et al., 2011). Sample sizes are increasing steadily and these collections now offer what has been previously unavailable to biological anthropologists: accurate, associated skeletal data that can be matched directly with known biological information for large numbers of individuals. One such sample provided the basis for this study.

1.5. Present research

As noted earlier, the three studies presented here aimed to address key issues relating to body mass estimation in biological anthropology by taking advantage of a large “virtual” collection of documented human subjects (Thali et al., 2007). Using volume-rendered skeletal models derived from CT scans and associated individual body masses, the first paper (Chapter 2) assessed three existing sets of regression equations for estimating body mass from cranial measurements. None of the equations had been tested on a population of known mass before, or been compared to each other. Accordingly, I tested each of the existing equations against a known-mass population

and directly compared their accuracy. This paper was published in the international peer-reviewed journal American Journal of Physical Anthropology (AJPA) in April 2014.

Paper 2 (Chapter 3) used the same sample as the first study, but focused on six sets of regression equations that have been developed for postcranial material. I evaluated the accuracy of these equations and also explicitly tested five assumptions relating to their relative performance. This paper was submitted to the journal Archaeological and Anthropological Sciences in December 2014 and is currently undergoing peer review. Publication is expected in the spring of 2015. The results of these first two papers suggest that existing equations should be used more cautiously than has been the practice in biological anthropology and inferences drawn from the resulting estimates may need to be re-examined.

The third paper of this dissertation (Chapter 4) sought to improve body mass estimation by rectifying three of the problems identified in the previous research: the use of small samples, indirect measures of key variables and unassociated skeletal and body mass data. Using part of same sample as the previous two studies as a training sample, I derived new regression equations for a suite of cranial and postcranial variables. The new regression equations were then evaluated for accuracy and reliability against a test sample of individuals drawn from the same population as the training sample. This provided the opportunity to compare body mass estimates made from cranial and postcranial material on the same known-mass individuals for the first time and allowed the exploration of several assumptions relating to how regression equations are evaluated for reliability. Due to a dearth of associated known-mass samples, this approach had not been taken before. This paper was also submitted to Archaeological and Anthropological Sciences in December 2014 and is under review.

Collectively, these studies sought to answer two questions: 1) how accurate are existing body mass estimate equations?; and 2) can we improve upon existing methods by deriving new equation from improved reference material? These are important goals because body mass estimates are integral to a wide range of conclusions that are made about the biology and behaviour of fossil and extant species. If existing equations are not accurate, or do not perform consistently, many of the inferences drawn from them must be reconsidered. If discrepancies are found, then extra caution must be taken

when applying such methods outside very limited circumstances. Similarly, if new equations cannot sufficiently improve accuracy or are only accurate for certain applications, then this must be considered in all future estimation attempts. While it is not likely that any one approach will estimate mass accurately in all contexts, whatever methods are employed, they must be demonstrated to be reliable or should not be used. In this regard, the results of the three studies described here have significant implications for palaeoanthropology, bioarchaeology and forensic anthropology.

Chapter 2.

Estimating fossil hominin body mass from cranial variables: An assessment using CT data from modern humans of known body mass

Authors: Marina Elliott, Helen Kurki, Darlene A. Weston and Mark Collard

Published: American Journal of Physical Anthropology (2014, 154:201-214)

Statement of Contributions of Joint Authors

Elliott, Marina C. (Candidate): Primary data collection, designing and establishing research methodology, data analysis and interpretation, writing and compilation of manuscript, preparation of figures and tables

Collard, Mark (Senior Supervisor): Supervising and assisting with research design, editing and co-author of manuscript

Kurki, Helen (committee member): assisting with research design, editing and co-author of manuscript

Weston, Darlene (committee member): assisting with research design, editing and co-author of manuscript

This Chapter is an exact copy of the journal paper referred to above.

2.1. Introduction

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, behavioural 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 et al., 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; McHenry, 1992; Aiello and Wood, 1994; Kappelman, 1996; Ruff et al., 1997; Spocter and Manger, 2007; Ruff et al., 2012). 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; McHenry, 1992; Ruff et al., 1997; 2012), 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; McHenry, 1992; Grine et al., 1995; Ruff et al., 1997; 2012) 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 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; Ruff, 2000; Auerbach and Ruff, 2004; Ruff et al., 2012; 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 goal of this study was to test the way in which cranial measurements are used to estimate fossil hominin body mass in palaeoanthropology. As the equations in question are derived from datasets consisting of means for multiple extant primate species, the assumption has been that they are not used to estimate 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.

2.2. Materials and Methods

2.2.1. 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, between18 and 90 years (M mean = 48.1 yrs, F mean = 51.2 yrs). 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 (years), body mass (kg) and stature (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 post-mortem 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 2.1 provides the summary data for the sample.

Table 2.1. Summary data for test sample

	Females (n=125)			Males (n=128)			Combined-sex (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 (yrs)	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[1]	25.1	6.4	14.3-46.5	25.8	4.6	15.4-46.9	25.4	5.6	14.3-46.9

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

2.2.2. Imaging and 3D 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, chest), are taken at 120 kV with mAs and field of view (FOV) adjusted for optimal resolution. Cranial data are reconstructed with slice thicknesses of 0.75 mm (0.375 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 re-measured according to the method outlined above. Differences between the measurements recorded on the physical and virtual skulls were less than 3% for all variables.

2.2.3. 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 2.2, Figure 2.1). Intra-observer repeatability for these measurements was tested by re-measuring them on nine randomly selected crania, with a three-week time lapse. Mean percent 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 x 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 = (π/4) x b x h). This followed Spocter and Manger (2007) who argued that it provided a more accurate reflection of true area than simple breadth x height. The third orbital area (ORBA3) employed a method similar to Kappelman (1996) in which two-dimensional 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 foramen magnum area (FMA1) was calculated as breadth x 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 Appendix (Table A1).

Table 2.2. Cranial variables

No.	Abbreviation	Description	References[1]
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 x h): product of breadth x height – in mm^2	AW[3], SM[11]
8	ORBA2	Orbital area (ellipse): calculated from breadth x 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 x h): product of breadth x height – in mm^2	AW[12], SM[3]
11	FMA2	Foramen magnum area (ellipse): calculated from breadth x height as an ellipse – in mm^2	SM[4]

1. 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]).

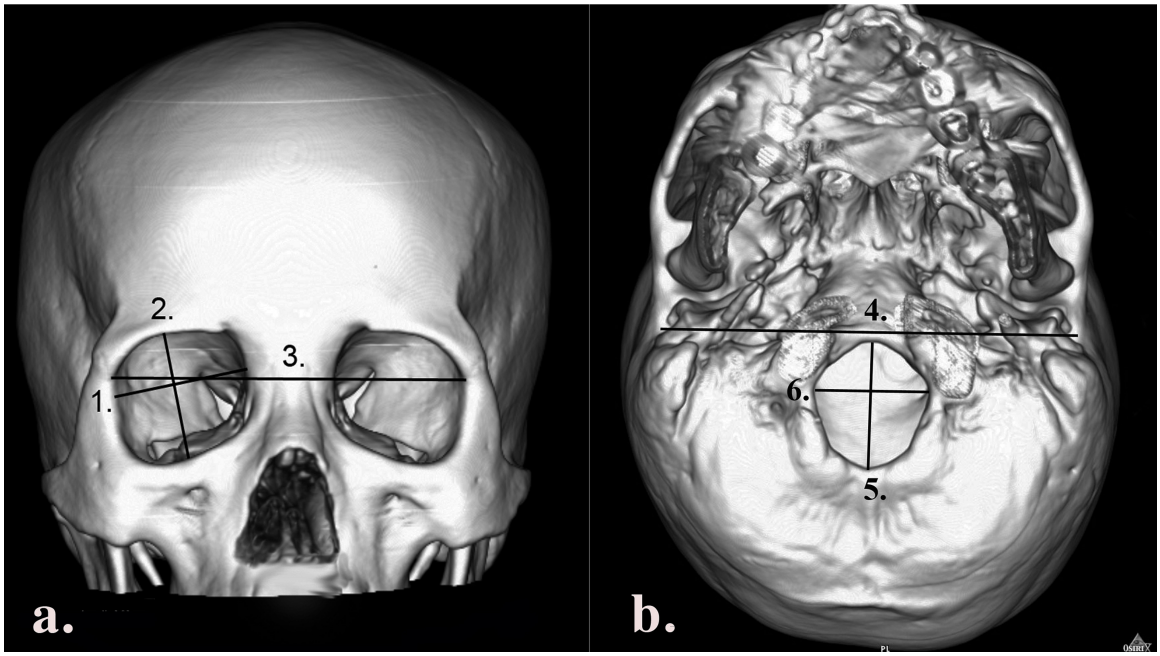


Figure 2.1. Frontal (a) and basicranial (b) example of skull three-dimensionally rendered from CT data, showing linear variables used for this study.

2.2.4. 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 2.3 lists the LSR-based equations tested, while Table 2.4 lists the RMA-based equations. For each set of equations, analyses were carried out on the full sample of 253 individuals, as well as sub-samples of females (n=125) and males (n=128).

Table 2.3. Published LSR equations for body mass estimation (hominoids)

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

Table 2.4. Published RMA equations for body mass estimation (hominoid)

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

All data log (base10) transformed. Measurements in mm or mm2, resulting BM in gm.

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; Konigsberg et al., 1998; Smith, 2009). Some researchers argue that Least Squares Regression (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; Konigsberg et al., 1998; Smith, 2009). 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 Reduced Major Axis (RMA) regression because it assumes that both variables are sampled with error and produces better results when extrapolating beyond the range of the original dataset (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 percent differences between the known and estimated body masses (EBM) were calculated for each individual. Raw differences were calculated as (known - EBM), while percent difference was calculated using the equation for percent 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 percent 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 estimated body mass fell within +/-20% of their known mass was also calculated. As a final comparison, we calculated the raw means (in kg) and 95% confidence intervals (CIs) for the predicted masses for each equation. However, for consistency with the

existing studies, the |PPE|s 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).

2.2.5. 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 percent differences between known and estimated body mass 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.

2.3. Results

The results of the tests of the LSR-based equations are summarized in Tables 2.5-2.7 and Figures 2.2-.2.4. Table 2.8 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 Appendix (Tables A2 and A3).

2.3.1. Aiello and Wood's (1994) equations

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

Table 2.5. Difference between known and estimated mass, Aiello and Wood (1994) LSR-based equations

	Females (n=125)			Males (n=128)			Combined-sex (n=253)		
Variable	PPE Mean[1] (SD)	|PPE| Mean (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean (SD)	20% (%)
BORB	11.4 (32.1)*	27.3 (20.3)	44.0	4.6 (30.2)*	25.1 (17.2)	42.2	8.0 (31.3)*	26.2 (18.8)	43.1
HORB	26.6 (28.9)*	34.4 (19.0)	28.0	36.8 (22.8)*	39.5 (17.8)	18.7	31.8 (26.4)*	36.9 (18.5)	23.3
BIOR	-10.4 (34.0)	25.0 (25.2)	56.8	-15.3 (30.5)*	24.2 (23.9)	53.9	-12.9 (32.3)*	24.6 (24.5)	55.3
BPOR	3.3 (28.2)*	21.4 (18.6)	57.6	**-1.4 (21.8)**	**17.5 (13.0)**	**64.8**	0.9 (25.2)*	19.4 (16.1)	61.3
LFM	-23.6 (39.6)*	33.5 (31.6)	41.6	-18.5 (34.3)*	28.1 (26.9)	52.3	-21.0 (37.1)*	30.8 (29.4)	47.0
BFM	-55.6 (60.9)*	62.6 (53.6)	25.6	-52.4 (54.1)*	56.7 (49.4)	21.9	-54.0 (57.5)*	59.6 (51.5)	23.7
ORBA1	22.2 (25.8)*	28.9 (18.0)	35.2	25.6 (20.8)*	28.7 (16.3)	29.7	23.9 (23.4)*	28.8 (17.1)	32.4
FMA1	-35.2 (45.4)*	41.9 (39.3)	36.0	-30.8 (37.9)*	36.0 (33.0)	37.5	-32.9 (41.8)*	38.9 (36.3)	36.8

PPE: percent prediction error (known - estimated)/known * 100, |PPE|: absolute percent prediction error, 20%: percent of individuals whose estimated body masses fall within +/-20% of known mass. 1. Directional differences (positive = underestimation, negative = overestimation); * indicates significance at p=0.01. Bold numbers indicate equations that met both acceptance criteria.

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.

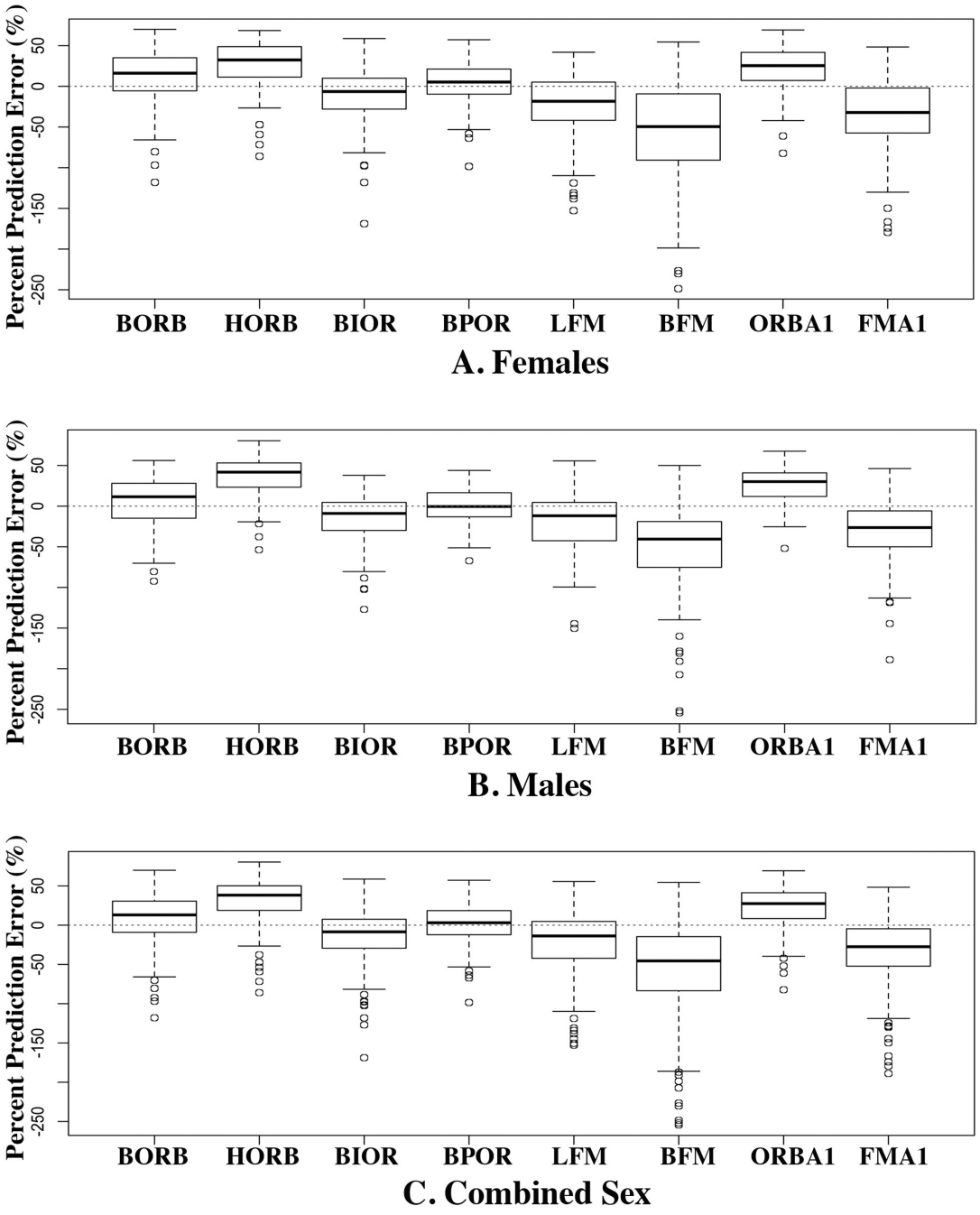


Figure 2.2. **Box plots of percentage of prediction error (PPE) between known and estimated masses for Aiello and Wood (1994) LSR-based equations: (a) females (n=125), (b) males (n=128), and (c) combined sex (n=253). Solid line = median, upper, and lower box margins = 75th and 25th percentiles, respectively, whiskers = limits of data still within 1.5 interquartile range (IQR) of Q1/Q3.**

2.3.2. 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.

Table 2.6. Difference between known and estimated mass, Kappelman (1996) LSR-based equations.

	Females (n=125)			Males (n=128)			Combined-sex (n=253)		
Variable	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)
HORB	24.4 (29.9)*	33.4 (19.2)	28.8	34.9 (23.5)*	38.1 (17.9)	21.9	29.7 (27.3)*	35.7 (18.9)	25.3
ORBA3	41.7 (18.9)*	42.8 (16.2)	8.0	45.1 (14.5)*	45.2 (14.2)	6.3	43.4 (16.9)*	44.0 (15.3)	7.1

PPE: percent prediction error (known - estimated)/known * 100, |PPE|: absolute percent prediction error, 20%: percent of individuals whose estimated body masses fall within +/-20% of known mass. 1. Directional differences (positive = underestimation, negative = overestimation); * indicates significance at p=0.01. Bold numbers indicate equations that met both acceptance criteria.

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.

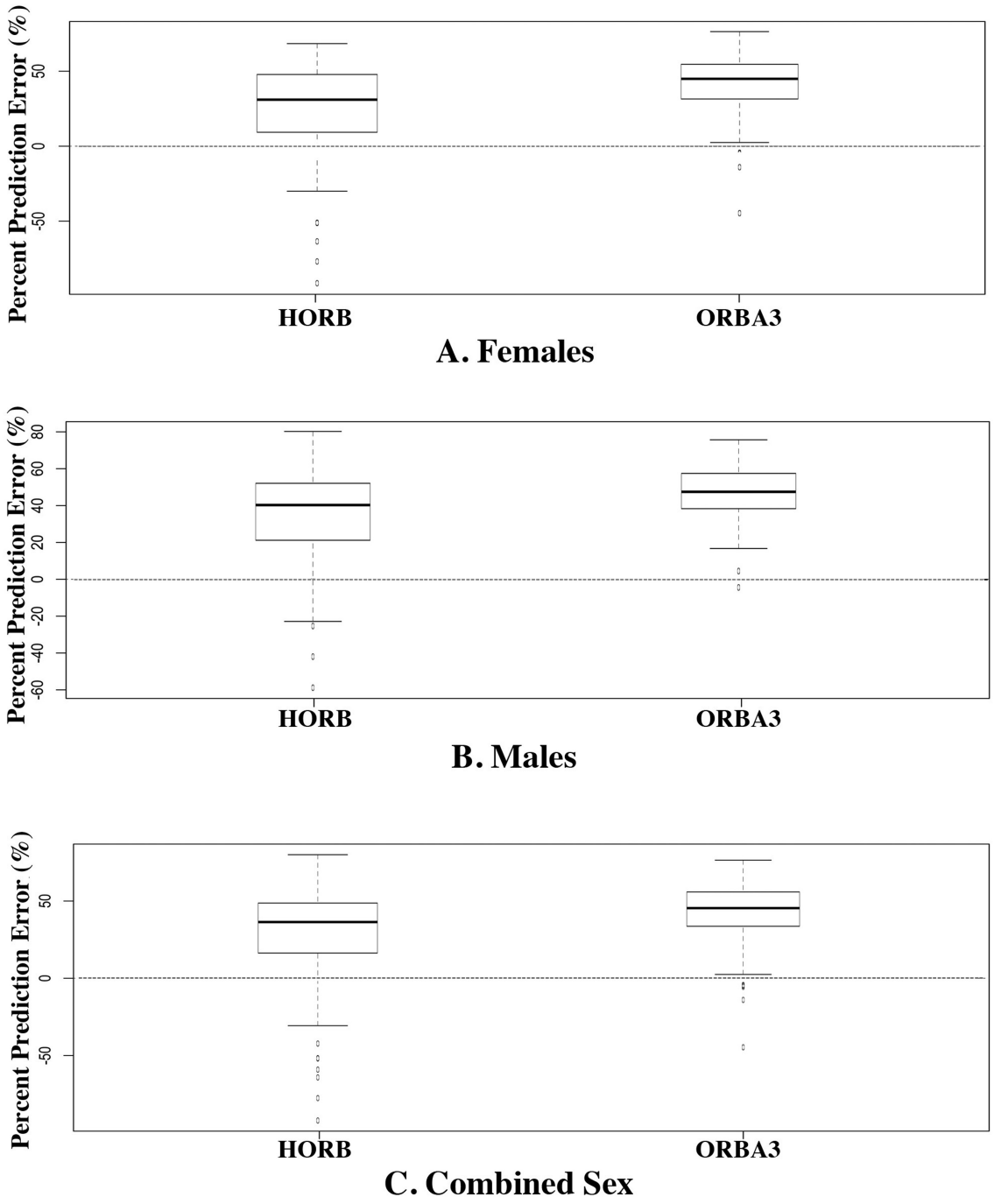


Figure 2.3. **Box plots of percentage of prediction error (PPE) between known and estimated masses for Kappelman (1996) LSR-based equations: (a) females (n=125), (b) males (n=128), and (c) combined sex (n=253). Solid line = median, upper, and lower box margins = 75th and 25th percentiles, respectively, whiskers = limits of data still within 1.5 interquartile range (IQR) of Q1/Q3.**

2.3.3. Spocter and Manger (2007) equations

Table 2.7. 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[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)
BORB	43.3 (17.9)	44.1 (15.8)	8.0	43.2 (15.0)	43.2 (14.8)	9.4	43.2 (16.5)*	43.6 (15.3)	8.7
HORB	75.9 (9.5)	76.0 (9.5)	0.0	79.3 (7.5)	79.3 (7.5)	0.0	77.6 (8.7)*	77.6 (8.7)	0.0
BIOR	73.6 (7.5)	73.6 (7.5)	0.0	73.7 (6.3)	73.7 (6.3)	0.0	73.7 (6.9)*	73.7 (6.9)	0.0
BPOR	-41.0 (44.8)	46.7 (38.7)	25.6	-57.8 (37.3)	58.2 (36.7)	15.6	-49.5 (42.0)*	52.5 (38.1)	20.5
LFM	28.9 (25.7)	34.8 (16.9)	19.2	29.2 (24.1)	33.1 (18.3)	23.4	29.1 (24.9)*	33.9 (17.6)	21.3
BFM	9.0 (35.8)	29.9 (21.5)	42.4	10.7 (31.9)	26.8 (20.4)	45.3	9.9 (33.8)*	28.3 (21.0)	43.9
ORBA1	61.6 (12.2)	61.6 (12.2)	0.1	63.9 (9.5)	63.9 (9.5)	0.0	62.8 (11.0)*	62.7 (11.0)	0.4
ORBA2	-21542.1 (9526.9)	21542.1 (9526.9)	0.0	-22827.0 (9304.4)	22863.0 (9304.4)	0.0	-22210.4 (9419.5)*	22210.4 (9419.5)	0.0
FMA1	16.2 (29.9)	28.2 (19.0)	38.4	17.1 (26.4)	26.9 (16.2)	39.1	16.7 (28.1)*	27.5 (17.6)	38.7
FMA2	-13924.5 (8485.9)	13924.5 (8485.9)	0.0	-16678.6 (10561.2)	16678.6 (10561.2)	0.0	-15317.9 (9672.1)*	15317.9 (9672.1)	0.0

PPE: percent prediction error (known - estimated)/known * 100, |PPE|: absolute percent prediction error, 20%: percent of individuals whose estimated body masses fall within +/-20% of known mass. 1. Directional differences (positive = underestimation, negative = overestimation); * indicates significance at p=0.01. Bold numbers indicate equations that met both acceptance criteria.

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 areas as ellipses produced extremely large prediction errors (>15,000%). It seems 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.

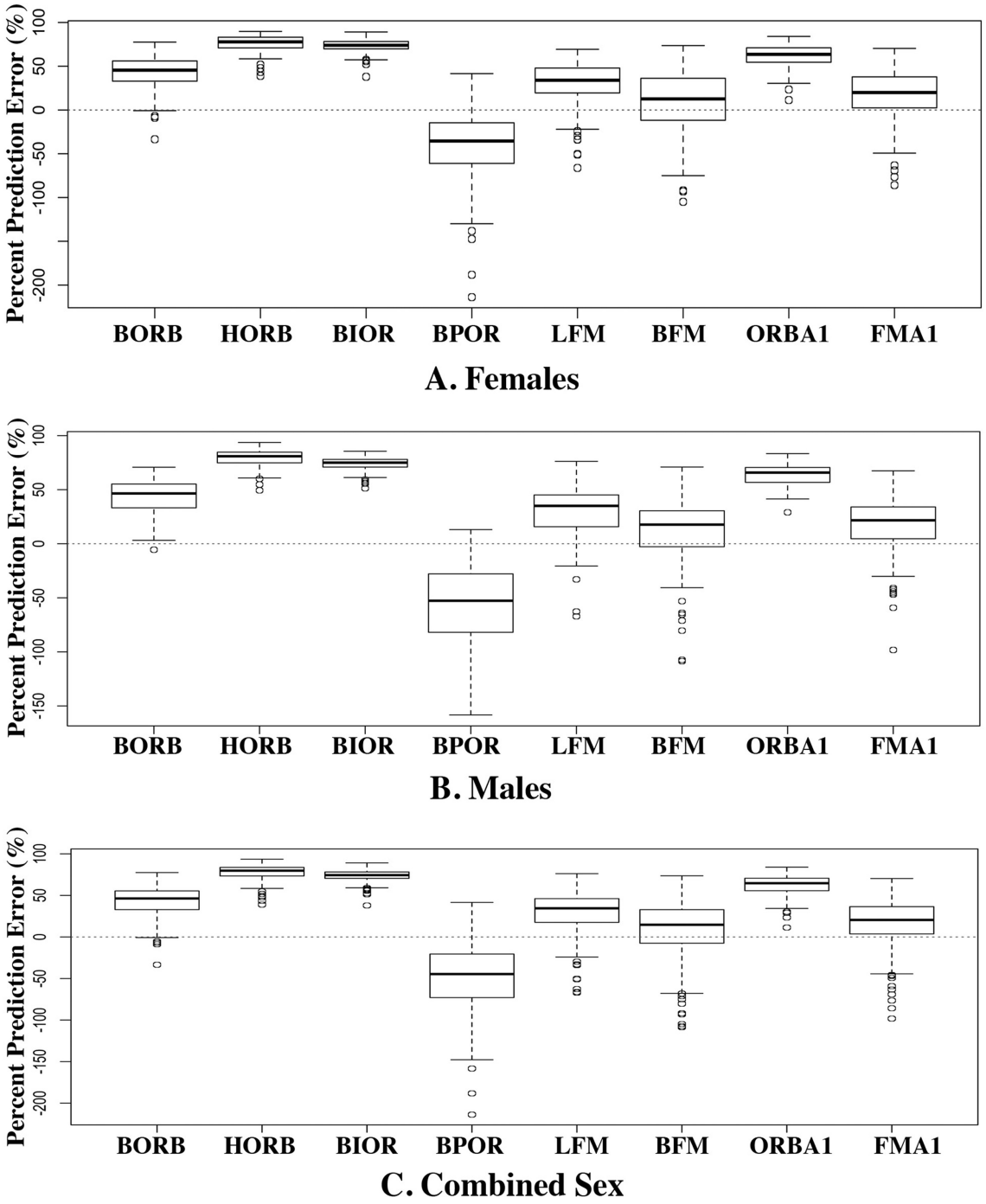


Figure 2.4. **Box plots of percentage of prediction error (PPE) between known and estimated masses for Spocter and Manger (2007) LSR-based equations: (a) females (n=125), (b) males (n=128), and (c) combined sex (n=253). Solid line = median, upper, and lower box margins = 75th and 25th percentiles, respectively, whiskers = limits of data still within 1.5 interquartile range (IQR) of Q1/Q3.**

2.3.4. Inter-study comparison of equations

Table 2.8. Comparison of mean body mass differences and confidence intervals for the combined sex sample (n=253), LSR-based equations

Variable	Source	predicted BM Mean (kg)	Difference from known mean* (kg)	95% C.I. for predicted mass
BORB	Aiello and Wood (1994)	66.5	9.1	64.1-68.9
	Spocter and Manger (2007)	40.8	34.8	39.8-41.9
HORB	Aiello and Wood (1994)	48.7	26.9	47.0-50.4
	Kappelman (1996)	50.2	25.4	48.4-51.9
	Spocter and Manger (2007)	16.0	59.7	15.4-16.5
BIOR	Aiello and Wood (1994)	81.7	6.1	79.4-84.0
	Spocter and Manger (2007)	19.0	56.6	18.6-19.4
BPOR	Aiello and Wood (1994)	71.8	3.8	70.0-73.6
	Spocter and Manger (2007)	109.0	33.3	105.4-112.5
LFM	Aiello and Wood (1994)	87.1	11.5	84.6-89.7
	Spocter and Manger (2007)	51.3	24.4	49.3-53.2
BFM	Aiello and Wood (1994)	110.9	35.3	106.5-115.4
	Spocter and Manger (2007)	64.9	10.7	62.3-67.6
ORBA1	Aiello and Wood (1994)	54.6	21.0	53.2-56.1
	Spocter and Manger (2007)	26.7	48.9	26.1-27.3
ORBA3	Kappelman (1996)	40.5	35.1	39.5-41.5
FMA1	Aiello and Wood (1994)	95.8	20.1	92.8-98.7
	Spocter and Manger (2007)	60.1	15.5	58.0-62.3

* Mean known body mass for full sample (n=253) is 75.62 kg.

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. 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 variable 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 19.0 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.

2.4. 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 (BPOR) 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 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 x 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, 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 (Appendix Table 4).

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 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 sub-sample of individuals (n=116) with BMIs in what the World Health Organization considers to be the "normal" range (18.5-25) (WHO, 2000). Appendix Tables 5 and 6 provide the sample and variable summaries for the BMI-restricted groups, while Appendix Tables 7-9 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 Appendix Table 10). 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,

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 LS regression are similar and LS regression 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 towards 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.

2.5. Conclusion

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.

Chapter 3.

Estimating body mass from postcranial variables: an evaluation of current equations using a large known-mass sample of modern humans

Authors: Marina Elliott, Helen Kurki, Darlene A. Weston and Mark Collard

Submitted: Archaeological and Anthropological Sciences (December 2014)

Statement of Contributions of Joint Authors

Elliott, Marina C. (Candidate): Primary data collection, designing and establishing research methodology, data analysis and interpretation, writing and compilation of manuscript, preparation of figures and tables

Collard, Mark (Senior Supervisor): Supervising and assisting with research design, editing and co-author of manuscript

Kurki, Helen (committee member): assisting with research design, editing and co-author of manuscript

Weston, Darlene (committee member): assisting with research design, editing and co-author of manuscript

This Chapter is an exact copy of the journal paper referred to above.

3.1. Introduction

Estimating body mass from skeletal remains is an important aspect of biological anthropology research. In palaeoanthropological and bioarchaeological contexts, body mass estimates offer one of the few ways to access key biological and behavioural information (Ruff, 2002; Plavcan, 2012). Estimates of mass also provide the most practical means for comparing individuals or groups, and relating these differences to their ecological and adaptive contexts (Smith, 1996). In addition, because body mass is a conspicuous individualizing feature and a potentially significant influence on taphonomic processes, it is becoming an area of interest for modern forensic identification and multiple fatality investigations (Rainwater et al., 2007; Agostini and Ross, 2011; Moore and Schaefer, 2011; Byard, 2012).

Body mass is most commonly estimated from the postcranium. Currently, there are two broad approaches - referred to as the “mechanical” and “morphometric” methods (Ruff, 2002; Auerbach and Ruff, 2004). Mechanical methods rely on the functional relationship between body mass and the skeletal elements that bear weight (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995). In contrast, morphometric methods estimate mass by reconstructing overall body shape from measures of stature and bi-iliac breadth (Ruff, 1994; Ruff et al., 1997; 2005). Although interspecies analyses have been used to estimate mass in fossil hominins (e.g. Steudel, 1980; McHenry, 1992; Hartwig-Scherer, 1993), the most commonly used methods are based on regression analyses that employ reference populations of modern humans, either individuals within a population (Ruff et al., 1991; 2012), or sex-specific means for multiple groups (McHenry, 1992; Ruff, 1994; Grine et al., 1995; Ruff et al., 2005).

The regression equations derived from these intraspecies studies have become the standard in biological anthropology. They have been used to estimate the mass of numerous fossil hominin groups (Ruff et al., 1997; Churchill et al., 2012), individual hominin specimens (Ruff and Walker, 1993; Ruff, 1994; Trinkaus and Jelinek, 1997; Arsuaga et al., 1999; Rightmire 2004; Rosenberg et al., 2006; Ruff et al., 2006; Melton et al., 2010; Ruff, 2010; Walker et al., 2011), and archaeological human populations (Kurki

et al., 2010; Myszka et al., 2012; Pomeroy and Stock, 2012; Ruff et al., 2012). The equations are also being used to explore body mass estimation as a tool in modern forensic contexts (Lorkiewicz-Muszyńska et al., 2013).

Despite their widespread use, a number of issues suggested that a re-evaluation of current postcranial equations was warranted. To begin with, the validity of the mechanical and morphological approaches has been based largely on the similarity of their results when compared to each other (Holliday, 2002; Auerbach and Ruff, 2004; Kurki et al., 2010; Myszka et al., 2012; Pomeroy and Stock, 2012; Ruff et al., 2012). However, there were still a number of discrepancies in the results of these studies and the similarities demonstrate only that the two sets of methods are roughly congruent, not that they are accurate.

In addition, the equations have not been adequately tested using populations of known mass, combined with associated skeletal data. For the mechanical method, only two studies have taken such an approach (Ruff et al., 1991; Lorkiewicz-Muszyńska et al., 2013). In the first study, Ruff et al. (1991) used skeletal measurements taken from clinical radiographs of living individuals (n= 80) and patient-recalled body masses to derive predictive equations for estimating body mass in modern and earlier human groups. However, in addition to using indirect measures for key variables, the study used a very small sample (n=8) to test the resulting equations. Consequently, it did not sufficiently demonstrate their accuracy.

The second study employed CT-derived skeletal measures and clinically documented body masses to test several sets of postcranial equations for their applicability to forensic identifications (Lorkiewicz-Muszyńska et al., 2013). Although documented body masses were used, the authors acknowledged a number of inconsistencies in the results. In particular, females were estimated more poorly than males and the STBIB equation performed worse than most of the FHB equations. Unfortunately, the reasons for the discrepancies were not fully explored. Nor were newly published mechanical equations (Ruff et al., 2012) included in the test. As a result, the predictive competence of the equations was not established.

The morphometric equations have also not been validated sufficiently. In particular, the most commonly used equations - those of Ruff et al. (1997) and Ruff et al.

(2005) – are both based on bi-iliac breadths measured over soft tissue on living individuals, combined with body masses gleaned from the literature. Again, since the equations were not derived using directly measured variables or associated material, the tests did not truly demonstrate the accuracy of the equations or their applicability to skeletal material.

Several other questions associated with the use of the equations also remain unanswered. The first is whether or not the biomechanical method is better than the morphometric method. Most researchers assume that the mechanical method is more accurate because the femur bears the majority of the body's weight (Jungers, 1988; Aiello and Wood, 1994; Churchill et al., 2012). However, the morphometric method has been argued to be more reliable than the mechanical method because it encompasses greater geographic diversity and uses larger sample sizes (Auerbach and Ruff, 2004). In fact, body masses estimated using the morphometric method are now being used as 'true masses' in studies deriving new predictive equations from archaeological material (Ruff et al., 2012). For this practice to be appropriate however, the reliability of the morphometric method must be explicitly demonstrated.

A second question relates to the choice of specific equation for the target specimen. Some authors have argued that variations in body size and proportion necessitate specific mechanical equations for smaller or larger bodied groups (McHenry, 1992; Grine et al., 1995; Kurki et al., 2010). If a specimen does not fit into one of these "target" groups, the use of "generalized" equations has been recommended (Ruff et al., 1991; 2012). Alternatively, Auerbach and Ruff (2004) recommend averaging the results of multiple mechanical equations for non-specific specimens. However, it is not always possible to know how well a specimen matches a reference sample (Smith, 2009). In addition, there is some question as to whether averaging estimates is statistically appropriate (Smith, 2002; SWGANTH, 2012). Either way, such claims must be clearly demonstrated before they can be accepted in practice.

In a similar vein, when choosing which morphometric equation to use, Ruff et al. (2005:390), argue that newer equations based on an expanded sample are "more broadly applicable, particularly to tall and wide-bodied males" than previous equations (Ruff, 1994; Ruff et al., 1997). As a result, the newer equations have largely replaced

earlier ones in practice (Ruff et al., 2006; 2012; Lorkiewicz-Muszyńska et al., 2013). However, the new equations were designed specifically to increase the representation of high-latitude populations in the sample (Ruff et al., 2005). Consequently, applying them uncritically to other test groups may not be appropriate and the assumption of their general superiority requires confirmation.

Lastly, the advantage of using sex-specific equations over combined-sex equations has not been clearly established. In general, sex-specific equations are considered to be more accurate than combined-sex equations because of systematic differences in body size between males and females (Ruff et al., 1991). However, sex is not always easy to attribute and the level of sexual dimorphism in the reference sample may not be the same as that of the target specimen (Niskanen and Junno, 2009). As a result, it has been suggested that combined-sex equations may be more accurate because they can employ larger reference samples and encompass a broader range of variation (Henneberg et al., 2005). Of the most commonly cited studies, two provide both sex-specific and combined-sex equations (Ruff et al., 1991; 2012). Two others provide only sex-specific equations (Ruff et al., 1997; 2005) and two provide only combined-sex equations (McHenry, 1992; Grine et al., 1995). In addition, Ruff (2000) recommends averaging the results of the male and female morphometric equations to create the equivalent of a 'combined-sex' equation for situations when sex cannot be assigned. Since the practice of averaging multiple estimates may be statistically problematic (Smith, 2002), assumptions regarding the relative predictive ability of equations designed to target specific groups must be formally evaluated with a known sample before they can be accepted.

Given the scope of the questions to which skeletally-derived body mass estimates are applied in biological anthropology, and the importance of the inferences drawn from their results, the validity of existing postcranial methods must be clearly demonstrated. Consequently, this study used virtually reconstructed skeletal elements derived from CT scans for a large sample of documented modern humans to systematically evaluate the predictive ability of the most widely used equations: four sets of "mechanical" (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995; Ruff et al., 2012), and two sets of "morphometric" (Ruff, 1994; Ruff et al., 1997; 2005) equations.

The accuracy rates from these tests were then used to evaluate five claims that have been made about the performance of the equations relative to one another: 1. Morphometric equations are more reliable than mechanical equations (Auerbach and Ruff, 2004); 2. “Matched-target” equations are more accurate than “generalized” equations and “mismatched-target” equation are less accurate than either (McHenry, 1992; Grine et al., 1995); 3. When using the mechanical method, if a specimen does not fit with one of the “target” equations, averaging the results of other equations will estimate mass reliably (Auerbach and Ruff, 2004); 4. Sex-specific equations are more accurate than those based on combined-sex samples (Ruff et al., 2012); and 5. When applying the morphometric method, if sex cannot be determined, averaging the sex-specific equations produces reliable estimates (Ruff, 2000).

3.2. Materials and Methods

3.2.1. 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., 2003, 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/stature2) 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 3.1 provides the summary data for the sample.

Table 3.1. Summary data for test sample

	Females (n=125)			Males (n=128)			Combined-sex (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 (yrs)	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[1]	25.1	6.4	14.3-46.5	25.8	4.6	15.4-46.9	25.4	5.6	14.3-46.9

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

3.2.2. Imaging and 3D reconstruction protocols

CT imaging was conducted according to IFM protocols (Thali et al., 2007). Patient scans were accessed from the IFM's archives and the regions of interest were volume rendered using OsiriX imaging software (http://www.osirix-viewer.com). The skeletal elements were oriented in a consistent plane and measured on the right side to the nearest 0.1 mm using OsiriX tools. In 13 cases where the right femur was unusable due to a fracture or prosthetic, the left side was measured on the grounds that directional asymmetry in the lower limbs is usually small and inconsequential for the purposes of estimating body mass (Auerbach and Ruff, 2004; Ruff et al., 2012). The accuracy of reconstructing virtual skeletal elements from CT data has been demonstrated for a number of applications (Cavalcanti et al., 2004; Lopes et al., 2008; Decker et al., 2011; Kim et al., 2012; Smyth et al., 2012). It was also verified in a previous study by physically measuring, scanning, virtually reconstructing, and then virtually re-measuring an archaeological skull from the IFM's collection (Elliott et al., 2014). In the latter study, measurement differences between the physical and virtual skulls were less than 3% for all variables.

3.2.3. Skeletal variables

Two skeletal measurements were taken for this study (Table 3.2, Figure 3.1): superior-inferior femoral head breadth (FHB) for use with the mechanical equations, and

bi-iliac breadth (BIB) for use with the morphometric equations. For the morphometric equations, stature (ST) was taken from patient documents, as measured by IFM staff at the time of processing. Intra-observer repeatability was tested by re-measuring both variables on six randomly selected individuals, with a three-week time lapse (mean percent errors < 1%). It is important to note that the morphometric equations use "living BIB", a measure that includes the cartilaginous soft tissue between the innominates (Ruff, 1994). When dealing with skeletal remains, Ruff (1994) recommends applying a conversion factor to skeletal BIB to obtain the "living" value needed for the equations. However, because the cartilage was still intact in the cadaveric individuals used here, this conversion was not necessary. Table 3.3 provides the summary data for the variables.

Table 3.2. Postcranial variables

No.	Abbreviation	Description	Reference
1	FHB	Femoral head breadth: superior-inferior breadth perpendicular to the cervical axis - in mm	Ruff et al. 1991
2	BIB	Bi-iliac breadth: maximum pelvic breadth taken across the iliac crests, taken as "Living BIB" - in mm	Ruff, 1991; 1994

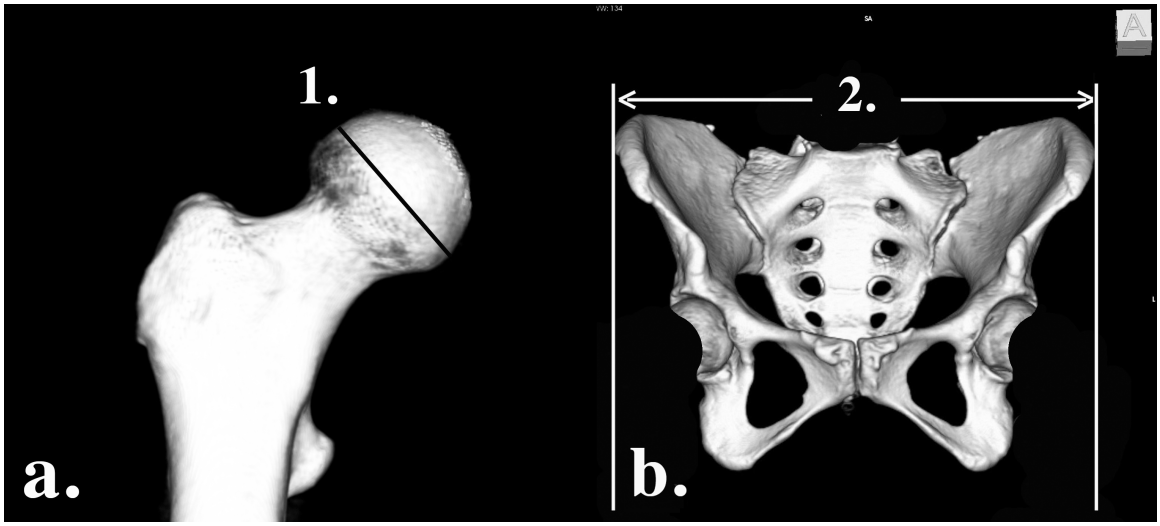


Figure 3.1. Post-cranial elements three-dimensionally rendered from CT data, showing linear measurements used for this study.

Table 3.3. Summary data for each variable (in mm)

	Females (n=125)			Males (n=128)			Combined-sex (n=253)		
Variable	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
FHB	45.5	2.3	39.8-55.5	50.9	2.8	41.5-57.5	48.2	3.73	39.8-57.5
BIB	277.9	18.5	223.1-337.1	283.9	16.8	211.2-324.2	280.9	17.83	211.2-337.1

3.2.4. Analyses

Table 3.4 lists the published equations tested in this study, along with the composition of the reference samples and the regression method used. Equations for the mechanical and morphometric equations are designated by the abbreviation for the variable/s used – FHB for the femoral head breadth-based equations and STBIB for the stature/bi-iliac breadth equations. Two studies provide sex-specific, as well as combined-sex, FHB equations (Ruff et al., 1991; 2012). Two others provide only combined-sex FHB equations (McHenry, 1992; Grine et al., 1995). Only sex-specific STBIB equations have been published (Ruff et al., 1997; 2005). Although there is considerable debate about the best regression method to employ for predictive analyses (Smith, 1996; Konigsberg et al., 1998; Smith, 2009; Sokal and Rohlf, 2012), most of the equations that have become standard were derived using Least-Squares Regression (LSR). The FHB-4 equations are exceptions to this. These equations were derived using Reduced Major Axis (RMA) regression (Ruff et al., 2012).

Analyses were conducted by entering the skeletal measurements into the appropriate equation and calculating an estimated mass. Raw and percent differences, percent prediction errors (PPE), and absolute percent differences (|PPE|) were calculated for each individual in three test groups: males, females and combined sexes. Raw differences were calculated as (known - estimated mass), while percent prediction errors were calculated as (known - estimated mass)/known*100 (Wu et al., 1995). PPEs indicate the directional difference of the error: positive PPE values indicate an underestimate (estimated mass < known mass), while negative values denote an overestimate (estimated mass > known mass). Absolute percent differences (|PPE|) assess the magnitude of the difference between the estimated and known masses (Dagosto and Terranova, 1992; Aiello and Wood, 1994). Differences between known and estimated mass were plotted and the Wilcoxon signed-rank test was used to

establish their significance. The percentage of individuals whose estimated body mass fell within +/-20% of the known mass was also calculated (see below). Assessment of reliability was based primarily on the |PPE| and the percentage of estimated masses that fell within +/-20% known mass. The |PPE|s were also used to compare the equations in relation to the assumptions previously discussed, with the Wilcoxon signed-rank test once again being used to determine the significance of the differences.

Table 3.4. Published regression equations for estimating body mass from femoral head breadth (FHB) or stature/bi-iliac breadth (STBIB)

Method	Female	Male	Combined-sex	Source	Sample composition and method
FHB-1	2.43*FHB-35.1	2.74*FHB-54.9	2.16*FHB-24.8	Ruff et al. 1991; 1997	80 living individuals (US whites and blacks), LS regression
FHB-2	n/a	n/a	2.24*FHB-39.9	McHenry, 1992, see Ruff et al. 1997	Mean data from 4 samples (US European and African, Khoisan and African Pygmy), LS regression
FHB-3	n/a	n/a	2.27*FHB-36.5	Grine et al. 1995	Mean data from 10 sex-specific samples (African American, European American and Native American), LS regression
FHB-4	2.18*FHB-35.8	2.80*FHB-66.7	2.30*FHB-41.7	Ruff et al. 2012	Archaeological sample of 1,145 individuals (European Holocene), RMA regression
STBIB-1	0.52*STAT+1.81*LBIB-75.50	0.37*STAT+3.03*LBIB-82.5	Average of male and female[1]	Ruff, 1994, Ruff et al. 1997; 2000	Sex-specific mean data for 56 samples (Worldwide), LS regression
STBIB-2	0.50*STAT+1.80*LBIB-72.60	0.42*STAT+3.13*LBIB-92.9	Average of male and female[1]	Ruff et al. 2005	Same data as for STBIB-1, with addition of 2 Finnish groups (1 male, 1 female), LS regression

All equations are for raw (non-logged) data. FHB in mm, stature (ST) and living bi-iliac breadth (LBIB) in cm. Resulting BM in kg. [1]as recommended by Ruff (2000).

3.2.5. Expectations

There are few clear criteria for the acceptance of a skeletally-derived body mass estimate. In theory, the predictive ability of a regression equation developed and tested on a single species (e.g. humans) should be higher than one that was developed on

multiple species and tested on one (Smith, 2002). Since studies using interspecies comparisons have considered percent prediction errors of up to 19% to be acceptable (Aiello and Wood, 1994), this would argue in favour of a lower threshold for the postcranial intraspecific equations being considered here. Despite this, because it provided a lenient baseline, we maintained this threshold and accepted estimates as reliable when the absolute percent prediction errors (|PPE|) fell below 19%. In addition, studies vary in terms of their expectations for the number of individuals that can be estimated with +/-20% of their known mass. Interspecific studies tend to use relatively high thresholds – accepting equations that place 50% of a sample within +/-20% of known mass (Dagosto and Terranova, 1992; Aiello and Wood, 1994). In contrast, intraspecific studies suggest that an effective equation should estimate the majority of individuals within 10 or 15% of their known mass (Ruff et al., 2005; Lorkiewicz-Muszyńska et al., 2013). Again, we chose a conservative approach and accepted estimates when 50% or more of the specimens fell within +/-20% of their known mass. In addition to these broad acceptance criteria, we made specific predictions for the performance of each equation, based on the original reference samples and intended purpose.

Mechanical (FHB) equations. Ruff et al.'s (1991) three FHB equations (FHB-1a-1c) were derived from individual data for a modern North American sample and were intended for general application to both fossil and modern humans. The reference sample is roughly contemporaneous with our test sample and the characteristics are similar (mean body mass=76.7 kg versus our 75.6 kg). In addition, Ruff et al. (1991) provide both sex-specific and combined-sex equations. As a result, we expected each of the equations to estimate mass accurately in their respective test groups.

McHenry's (1992) single FHB equation (FHB-2) was derived from mean, combined-sex data for four modern samples (US European and African, Khoisan and African "Pygmy"). Because it was designed specifically to estimate mass in smaller-bodied hominins, this equation was not expected to be accurate when applied to our sample, which consists of relatively large, modern Europeans. In light of this mismatch, FHB-2 was also expected to estimate mass relatively more poorly than the other equations.

The combined-sex FHB equation provided by Grine et al. (1995) (FHB-3) was derived from mean data for ten sex-specific modern and archaeological samples, with the goal of estimating large-bodied hominins. Initially, this suggested to us that the equation might overestimate mass in the current sample, as Pleistocene hominins are considered to have been heavier and more robust than modern humans (Ruff et al., 2005; Churchill et al., 2012). However, as a Western, industrialized, modern group, our sample might also be carrying more fat than the original reference populations (Ruff et al., 1991) and the equation would underestimate mass. On the grounds that these two factors would effectively cancel each other out (Ruff, 2000), we expected the FHB-3 equation to predict mass well here.

Ruff et al.'s (2012) recently derived FHB equations (FHB-4a-4c) provide sex-specific, as well as combined-sex equations for Holocene European samples. Because they were derived from a large and diverse group of individual skeletal remains, these equations have been argued to be "broadly applicable across different geographic regions and temporal periods" (Ruff et al., 2012:615). Following this, the equations were expected to estimate mass well in our sample.

Morphometric (STBIB) equations. The two sex-specific STBIB-1 equations tested here were developed using population-mean anthropometric data taken from living individuals belonging to 56 different populations (Ruff, 1994; Ruff et al., 1997). As associated data were not available, mean body masses for each group were gleaned from the literature (Ruff, 1991). Designed to encompass a wide range of body sizes, these equations have been argued to "provide the most generally reliable" estimates of mass when bi-iliac breadth and stature can be measured or estimated with some confidence (Auerbach and Ruff, 2004:340). As we were able to measure both features directly in our study sample, we expected the equations to perform well.

As noted earlier, the STBIB-2 equations are the result of efforts to improve body mass estimates by expanding the range of variation in the reference sample to include more high-latitude (> 46º N) populations (Ruff et al., 2005). As another high-latitude, tall and broad population (Zurich is at 47.4° N and all means either met or exceeded those for the Inupiat and Finnish groups), we expected body mass to be estimated well with these equations, particularly for the males.

3.2.6. Relative performance

We also tested specific expectations for the relative performance of the equations based on the claims previously discussed. The first hypothesis states that morphometric equations are more reliable than mechanical ones (Auerbach and Ruff, 2004). Thus, we expected STBIB-1 and STBIB-2 to outperform all of the FHB equations.

According to the second hypothesis, "matched-target" equations are more accurate than "generalized" or "mismatched-target" equations. Here, we expected our modern European sample to be estimated better using the large-bodied FHB-3 equation than the more "generalized" FHB-1 and FHB-4 equations (at least with the combined-sex equations). However, all three of these equations were expected to perform better than the "mismatched" FHB-2 equation, which was designed for small-bodied individuals. On the same grounds, because our test sample was derived from a relatively high-latitude population, we expected the "matched-target" STBIB-2 equation to be more accurate than the more "generalized" STBIB-1 equation, particularly for men.

A third hypothesis claims that the results of multiple mechanical equations can be averaged when a specimen is not specifically large- or small-bodied (Auerbach and Ruff, 2004). Thus, we expected the average of FHB-1, FHB-2 and FHB-3 to estimate mass well, and at least as well as the generalized equations. We also expected the average of four mechanical equations (FHB-1-4) to perform well.

The fourth hypothesis assumes that sex-specific equations will perform better than combined-sex equations (Ruff et al., 2012). As a result, we expected FHB-1, FHB-4, STBIB-1 and STBIB-2 to return lower error rates than FHB-2 or FHB-3.

Lastly, the fifth hypothesis argues that it is appropriate to average the results of the sex-specific equations if sex is uncertain (Ruff, 2000). Although we still expected males and females to be estimated best with their respective sex-specific morphometric equations, averaging the two results was expected to produce reliable estimates.

3.3. Results

Table 3.5 and Figure 3.2 summarize the results for each of the FHB and STBIB equations. Table 3.6 provides the raw mean predicted masses, difference from known mean and 95% confidence intervals for the predicted masses.

Table 3.5. Differences between known and estimated body masses for each equation.

	Female (n=125)			Male (n=128)			Combined-sex (n=253)		
Method	PPE Mean[1] (SD)	|PPE| Mean (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean (SD)	20% (%)
FHB-1	-15.5 (29.7)	25.4 (21.7)	48.0	-7.2 (20.2)	**16.1 (14.2)**	**71.1**	-10.4 (25.0)	20.2 (18.1)	59.7
FHB-2	5.0 (24.6)	19.9 (15.2)	58.4	6.1 (17.7)	**15.0 (11.2)**	**73.4**	5.5 (21.3)	**17.4 (13.5)**	**66.0**
FHB-3	-2.3 (26.4)	20.7 (16.4)	56.0	-0.1 (18.9)	**14.7 (11.9)**	**75.8**	-1.2 (22.9)	**17.6 (14.6)**	**66.0**
FHB-4	2.8 (25.1)	19.9 (15.4)	57.6	4.0 (18.1)	**14.7 (11.3)**	**72.7**	4.0 (21.7)	**17.2 (13.7)**	**66.4**
STBIB-1[a]	6.2 (22.1)	**18.6 (13.3)**	**64.8**	11.8 (15.7)	**16.1 (11.3)**	**66.4**	8.6 (19.7)	**17.5 (12.4)**	**62.9**
STBIB-2[a]	6.5 (22.0)	**18.6 (13.4)**	**63.2**	10.7 (15.8)	**15.4 (11.2)**	**68.8**	9.1 (19.6)	**17.6 (12.4)**	**62.1**

PPE: percent prediction error (known - estimated)/known * 100, |PPE|: absolute percent prediction error, 20%: percent of individuals whose estimated body masses fall within +/-20% of known mass. 1 Directional differences (positive values indicate underestimation, negative values indicate overestimation); a) combined-sex values are the average of the male and female estimates as recommended by Ruff (2000). Bold values indicate the analyses that achieved mean |PPE|s below 19% and estimated more than 50% of the sample within +/-20% of known mass.

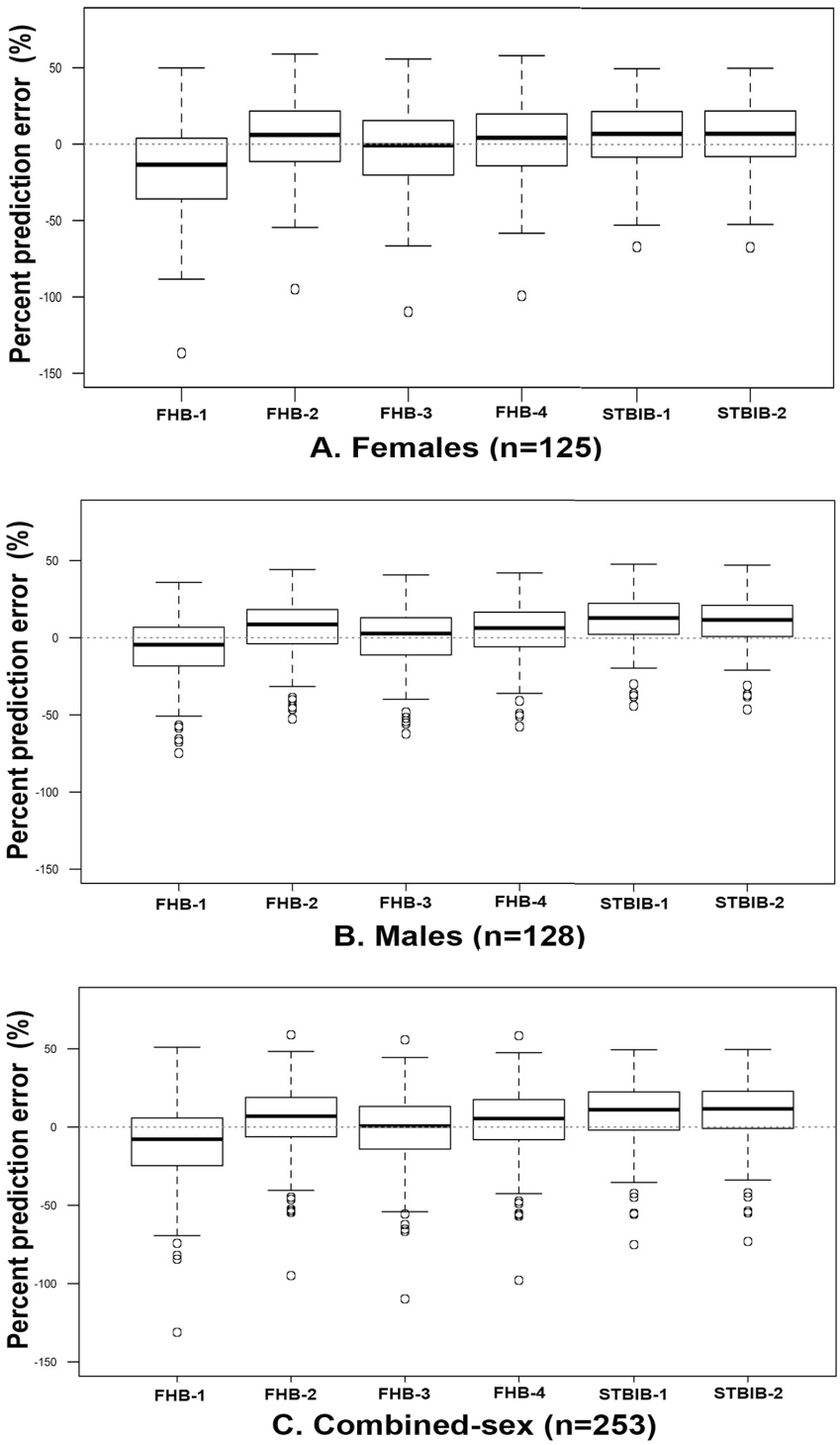


Figure 3.2. Box plots of percent prediction error (PPE) between known and estimated masses for post-cranial equations. Solid line=median, upper and lower box margins=75th and 25th percentiles respectively, whiskers=limits of data still within 1.5 IQR of Q1/Q3.

Table 3.6. Predicted mass (kg), mean difference (kg) and confidence intervals for each test sample.

	Female (n=125)			Male (n=128)			Combined-sex (n=253)		
Method	Predicted mean (SD)	Raw diff.[1] mean (SD)	95% CI	Predicted mean (SD)	Raw diff.[1] mean (SD)	95% CI	Predicted mean (SD)	Raw diff.[1] mean (SD)	95% CI
Known mass	69.5 kg	--	--	81.6 kg	--	--	75.6 kg	--	--
FHB-1	75.2 (5.4)	-5.7 (19.3)*	74.3-76.2	84.6 (7.7)	-3.1 (15.0)*	83.3-86.0	79.4 (8.1)	-3.7 (17.2)*	78.4-80.4
FHB-2	61.9 (5.1)	7.6 (19.3)*	61.0-62.8	74. 1 (6.3)	7.5 (14.9)*	73.0-75.2	68.1 (8.4)	7.6 (17.2)*	67.0-69.1
FHB-3	66.6 (5.2)	2.9 (19.3)	65.7-67.6	79.0 (6.3)	2.6 (14.9)	77.9-80.1	72.9 (8.5)	2.8 (17.2)	71.8-73.9
FHB-4	63.3 (5.0)	6.2 (19.2)*	62.5-64.2	75.8 (7.8)	5.7 (15.0)*	74.5-77.2	69.2 (8.6)	6.4 (17.2)*	68.2-70.3
STBIB-1	61.6 (6.0)	8.0 (17.7)*	60.5-63.6	69.8 (6.8)	11.8 (14.4)*	68.6-71.0	66.0 (7.2)	9.6 (16.2)*	65.1-66.8
STBIB-2	61.3 (5.9)	8.2 (17.7)*	60.3-62.4	70.7 (7.3)	10.8 (14.3)*	69.5-72.0	65.6 (7.3)	10.0 (16.2)*	64.7-66.5

[1] Positive values indicate that the predictive equation underestimates true mass, negative values indicate that the equation overestimates true mass. * indicates predicted vs. known mass difference significance at p=0.05.

3.3.1. Mechanical (FHB) equations

Using Ruff et al.'s (1991) sex-specific FHB-1a and FHB-1b equations respectively, the male sub-sample was estimated within acceptable limits, but the female sub-sample was not. For males, the mean |PPE| was below 16.1% and 71.1% of the individuals were estimated within +/-20% of their known mass. In contrast, for females the mean |PPE| exceeded 19% and less than half the sample (48%) was estimated within +/-20% of known mass. The combined-sex equation (FHB-1c) only partially met the acceptance criteria. More than 50% (59.7%) of the individuals were estimated within +/-20% of known mass, but the mean |PPE| exceeded 19% (20.2%). In terms of the direction of error, all three equations overestimated mass on average in their respective samples. Thus, the FHB-1 equations did not consistently estimate mass within acceptable limits in the test samples.

McHenry's (1992) single, combined-sex FHB-2 equation estimated mass within acceptable limits in the male and combined-sex test groups. In both cases, mean |PPE|s

were below 19% and more than 50% of the individuals were estimated within +/-20% of their known mass. In contrast, 58.4% of the female sample was estimated within +/-20% of known mass, but the |PPE| exceeded 19% (albeit by a small margin at 19.9%). The direction of error was again consistent across the three test groups, but masses were underestimated. Thus, this equation also failed to estimate mass reliably in all test groups.

Grine et al.'s (1995) equation (FHB-3) also resulted in mean estimates that met both criteria for acceptance in the male and combined-sex samples. Males showed the lowest mean error and estimated the highest number of individuals within +/-20% of known mass (>75%). However, only one of the two acceptance criteria was met in the female sample (more than 50% were estimated within +/-20% of known mass, but the |PPE| exceeded 19%). In terms of directional error, FHB-3 overestimated mass on average in all three groups. This equation did not estimate mass consistently, or within acceptable limits, across the three test samples.

Of Ruff et al.'s (2012) three sample-specific equations, the ones for males (FHB-4b) and combined-sexes (FHB-4c) met the criteria for acceptance. However, the female-only equation (FHB-4a) failed to estimate the test sample within the 19% |PPE| criterion (again, by a small margin: 19.9%). Directionally, the three equations all underestimated mass in their respective test groups. Thus, only two of the three equations estimated mass in the test samples within acceptable limits.

3.3.2. Morphometric (STBIB) equations

Ruff et al.'s (1997) two sex-specific equations (STBIB-1a and 1b) resulted in estimates that met the criteria for acceptance in their respective test groups. Ruff et al.'s (2005) sex-specific equations (STBIB-2a and 2b) also resulted in estimates that fell within acceptable limits for each test group. All four equations underestimated mass on average. Thus, both sets of equations estimated mass reliably in the test groups.

3.3.3. Hypotheses

1. *The morphometric/STBIB method estimates body mass more reliably than the mechanical/FHB method.* Table 3.7 summarizes the results for this comparison. As can

be seen, not all the results are consistent with the prediction. In females, the STBIB-1a and STBIB-2a equations estimated mass more reliably than most of the FHB equations, to a statistically significant level ($p<0.05$). However, neither was significantly better than the FHB-4 equations. In males, neither the STBIB-1b nor the STBIB-2b equation estimated mass more reliably than the FHB equations.

Table 3.7. Expectation that morphometric (STBIB) equations will achieve greater accuracy than the mechanical (FHB) equations

Expectation	Female (n=125)		Male (n=128)		Combined-sex (n=253)	
	P value	Expectation [1] met?	P value	Expectation [1] met?	P value	Expectation[1] met?
STBIB-1 more accurate than FHB-1	0.00*	Y	0.95	=	0.02*	Y
STBIB-1 more accurate than FHB-2	0.05*	Y	0.12	N	0.78	N
STBIB-1 more accurate than FHB-3	0.04*	Y	0.17	N	0.65	Y
STBIB-1 more accurate than FHB-4	0.08	Y	0.10	N	0.57	N
STBIB-2 more accurate than FHB-1	0.00*	Y	0.65	Y	0.03*	Y
STBIB-2 more accurate than FHB-2	0.05*	Y	0.51	N	0.62	N
STBIB-2 more accurate than FHB-3	0.04*	Y	0.44	N	0.98	=
STBIB-2 more accurate than FHB-4	0.08	Y	0.36	N	0.44	N

[1] based on |PPE| values, "=" indicates |PPE|s were the same for both equations; * indicates the difference between methods is significant ($p=0.05$)

The results for the sample as a whole were mixed. As noted previously, because separate equations were not provided, the combined-sex STBIB-1c or 2c results are the product of averaging the male and female estimates. However, there is still value in considering these results in the context of comparing the mechanical and morphometric methods. Here, the STBIB-1c or STBIB-2c averages were only more reliable than the FHB-1c equation, at a significant level. The remaining combinations either went against the prediction, or were statistically insignificant. Overall, the claim that the morphometric method produces more reliable estimates than the mechanical method was not consistently supported.

2. *"Matched-target" equations are more accurate than "generalized" equations and "mismatched-target" equations are less accurate than either the "matched-target" or*

"generalized" equations. The results of this test are summarized in Table 3.8. This claim was only partially supported in the test sample. In females, three equations performed as expected. However, one matched-target equation (FHB-3) went against the prediction and was significantly less accurate than a generalized equation (STBIB-1). The expectation that the mismatched FHB-2 equation would estimate mass less accurately than the matched equations, held only in relation to STBIB-2 and not to FHB-3. The mismatched FHB-2 equation estimated mass significantly worse than the generalized STBIB-1 equation, and went against expectations by estimating mass significantly better than the generalized FHB-1 equation. In the male sample, two matched-target equations met expectations and estimated mass significantly more accurately than the general equations. However, none of the other combinations met the prediction to statistically significant levels. For the combined-sex sample, the matched-target FHB-3 equation and STBIB-2 average were significantly more accurate than the general FHB-1 equation. Against expectations, the general FHB-1 equation was not more accurate than the mismatched FHB-2 equation in this group. Thus, the results do not support the assertion that body mass will be estimated most reliably by using a "matched-target" equation.

3. *When using the mechanical method, if a specimen does not fit with one of the "target" equations, taking the average of the results from other equations produces reliable estimates.* Table 3.9 summarizes these results. This hypothesis was also partially supported. In males, both the three and four-average estimates resulted in final estimates that met the criteria for acceptance. In this group, the averaged results estimated mass as well as, or better than, using a single equation. The same pattern held for the combined-sex samples – both the average of three and the average of four estimates produced mean estimates that met the criteria for acceptance, and performance was better than most of the single FHB equations. In females, averaging the results of three (FHB-1-3) or four (FHB-1-4) estimates did not result in a final estimate that met the acceptance criteria for reliability. However, the estimation accuracy was not made worse by averaging estimates in this group. Using the average of four equations was better than using the average of three equations in the female and combined sex samples ($p<0.05$), but in not the male group. These results partially support the practice of averaging different equations for a "generalized" specimen.

Table 3.8. Expectation that "matched-target" equations will achieve greater accuracy than "generalized" equations, but that the "mismatched-target" equation will achieve lower accuracy than the "generalized" equations

Expectation	Female (n=125)		Male (n=128)		Combined-sex (n=253)	
Matched target vs. generalized	P value	Expectation[1] met?	P value	Expectation[1] met?	P value	Expectation[1] met?
FHB-3 more accurate than FHB-1	0.00*	Y	0.02*	Y	0.00*	Y
FHB-3 more accurate than FHB-2	0.22	N	0.56	Y	0.58	N
FHB-3 more accurate than FHB-4	0.08	N	0.93	=	0.22	N
FHB-3 more accurate than STBIB-1	0.04*	N	0.17	Y	0.65	N
STBIB-2 more accurate than FHB-1	0.00*	Y	0.65	Y	0.03*	Y
STBIB-2 more accurate than FHB-2	0.05*	Y	0.51	N	0.62	N
STBIB-2 more accurate than FHB-4	0.08	Y	0.36	N	0.44	N
STBIB-2 more accurate than STBIB-1	0.57	=	0.00*	Y	0.00*	N
Mismatched target vs. generalized						
FHB-2 less accurate than FHB-1	0.00*	N	0.31	N	0.00*	N
FHB-2 less accurate than FHB-4	0.95	=	0.26	Y	0.09	Y
FHB-2 less accurate than STBIB-1	0.05*	Y	0.12	N	0.78	N

[1] based on |PPE| values, "=" indicates |PPE|s were the same for both equations; * indicates difference between methods is significant (p=0.05)

Table 3.9. Differences between known and estimated body masses when multiple mechanical methods are averaged

	Female (n=125)			Male (n=128)			Combined-sex (n=253)		
Method	PPE Mean[1] (SD)	|PPE| Mean (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean (SD)	20% (%)
Mean of FHB 1-3[a]	-4.3 (26.9)	21.1 (17.0)	58.4	-0.4 (18.9)	**14.7 (11.9)**	**75.8**	-2.0 (23.1)	**17.8 (14.8)**	**66.4**
Mean of FHB 1-4[b]	-2.5 (26.4)	20.7 (16.5)	56.8	0.7 (18.7)	**14.6 (11.7)**	**75.0**	-0.5 (22.7)	**17.5 (14.4)**	**66.4**

[1] Directional differences (positive values indicate underestimation, negative values indicate overestimation); 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) as recommended in Auerbach and Ruff (2004); b) as calculated in this study.

4. *Sex-specific equations are more accurate than the combined-sex equations. Table 3.10 summarizes these results*. This hypothesis was also only partially supported. In males, most of the comparisons went against expectations, but the differences were not statistically significant. The differences between FHB-4 and the two combined-sex equations were also not statistically significant. However, in females, the sex-specific FHB-1 equation estimated mass significantly worse than the combined-sex FHB-2 or FHB-3 equations ($p<0.05$). In the context of this assertion, the sex-specific STBIB-1 and STBIB-2 equations were also expected to estimate mass better than the combined-sex FHB-2 and FHB-3 equations. This was true for the female test sample (all statistically significant at $p<0.05$), but not for the males. Thus, sex-specific equations are not necessarily more reliable than combined-sex equations.

Table 3.10. Expectation that sex-specific equations will achieve greater accuracy than mixed-sex equations

Expectation	Female (n=125)		Male (n=128)	
	P value	Expectation[1] met?	P value	Expectation[1] met?
FHB-1 more accurate than FHB-2	0.00*	N	0.31	N
FHB-1 more accurate than FHB-3	0.00*	N	0.02*	N
FHB-4 more accurate than FHB-2	0.95	=	0.26	Y
FHB-4 more accurate than FHB-3	0.08	Y	0.93	=
STBIB-1 more accurate than FHB-2	0.05*	Y	0.12	N
STBIB-1 more accurate than FHB-3	0.04*	Y	0.17	N
STBIB-2 more accurate than FHB-2	0.05*	Y	0.51	N
STBIB-2 more accurate than FHB-3	0.04*	Y	0.44	N

[1] based on |PPE| values, "=" indicates |PPE|s were the same for both equations; * indicates difference between methods is significant (p=0.05)

5. *When applying the morphometric method, if sex cannot be determined, mass will be estimated reliably by taking the average of the sex-specific equations*. The results, summarized in Table 3.5, support this hypothesis. For both STBIB-1 and STBIB-2, taking the mean of the male and female results produced estimates that met both criteria for acceptance. Not surprisingly, using the sex-averaged equations resulted in error rates that fell between those of the two sex groups, and overall, did not significantly reduce accuracy. In this context, the results support the hypothesis that averaging the results of the sex-specific equations will estimate mass reliably.

3.4. Discussion

Many of the equations for estimating body mass from postcranial material met the criteria for acceptance in the male and the combined-sex samples. However, this was not the case for the combined-sex FHB-1 equation. In addition, none of the mechanical/FHB equations estimated mass reliably in the female sample. The equations also did not consistently perform as expected given their reference samples and target groups. For example, Ruff et al.'s (1991) FHB-1 equations were the least accurate estimators of mass in the test sample. This was despite being derived from modern individuals with characteristics similar to those of the test sample and including sex-specific equations. The FHB-4 equations did not estimate mass as well as expected despite being derived from a much larger sample, providing sex-specific equations and being described as "superior" to all other methods (Ruff et al., 2012:601). In contrast, McHenry's (1992) FHB-2 equation estimated mass better than anticipated given the single combined-sex equation and "mismatched" test and reference samples. Even with the lenient criteria used in the present study, none of the FHB equations estimated mass to the level of acceptance for reliability in females.

We also found mixed support for the claims that have been made regarding the way the equations should perform relative to one another. The morphometric/STBIB equations did not estimate mass more reliably than the mechanical/FHB methods in all groups. Females were estimated better with the STBIB equations, but males were estimated better with the FHB equations. When the sexes were combined, the morphometric/STBIB equations estimated mass less reliably than all but one of the mechanical/FHB equations (FHB-1c). In light of these results, it is not appropriate to use the morphometric/STBIB equations in preference to the mechanical/FHB equations.

Similarly, using "matched-target" equations did not consistently improve estimation accuracy over "generalized" equations. In keeping with the predictions, the FHB-3 (Grine et al., 1995) equation designed for large-bodied hominins estimated mass better than the general FHB-1 equation (Ruff et al., 1991) in the combined-sex group. However, it did not perform as well as the other general FHB-4 equation (Ruff et al., 2012). McHenry's (1992) FHB-2 equation met expectations and estimated mass less reliably than the general FHB-4 equation. However, as noted previously, it performed

better than the generalized FHB-1 equation despite being designed to estimate mass in small-bodied hominins. It also unexpectedly estimated males better than it did females. For the morphometric method, the "matched-target" STBIB-2 equation (Ruff et al., 2005) estimated mass more reliably than the more general STBIB-1 equation (Ruff et al., 1997) in males. However, it did not estimate mass better in the female test sample or when the sexes were combined. Although the results support Ruff et al.'s (2005) suggestion that the newer equations may be more appropriate for estimating large, high-latitude males, they also suggest that broadening the reference sample does not necessarily provide better estimation and the STBIB-2 equations should not simply replace the STBIB-1 equations uncritically.

In partial support of the claim that averaging the results of multiple equations produces reliable body masses (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995), the male and combined-sex groups were estimated slightly more accurately using the average of three FHB equations than using a single equation. In females, although the error rates decreased slightly compared to most of the single FHB equations, averaging the results of FHB-1a, FHB-2a and FHB-3a still did not produce mean estimates that met the criteria for acceptance. Averaging the results of four FHB equations (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995; Ruff et al., 2012) also resulted in a modest improvement compared to using a single equation in the male and combined-sex samples. But again, the four-estimate average failed to estimate females to an acceptable level. In general, averaging multiple FHB equations has a neutral, or slightly positive effect on predictive accuracy.

Sex-specific equations were not consistently more accurate than those designed for combined sexes. The sex-specific FHB-1 or FHB-4 equations should have achieved greater accuracy than either of the combined-sex FHB equations (McHenry, 1992; Grine et al., 1995). However, this was not the case, particularly in females. This result suggests that the use of sex-specific equations may not be critical, at least in species like humans, who exhibit relatively low levels of sexual dimorphism (Ruff, 2002). In fact, it may be that the greater number of individuals in the combined-sex reference group is driving the improved accuracy and that large sample sizes are more important than group-specificity for developing predictive equations.

Lastly, averaging the male and female morphometric/STBIB estimates in the absence of known sex, produced results that met the acceptance criteria. |PPE|s for the averaged values were slightly lower than those for females, but higher than for males. Fewer individuals were estimated within +/-20% of known mass using the average of the sexes than either sex alone. However, in both cases, the differences were small. The results of this test suggest that there does not appear to be a significant cost to accuracy by averaging sex-specific equations.

Several of these results require further consideration. First, although they were surprising in relation to the claims in the literature, there may be a simple explanation for the relatively poor performance of Ruff et al.'s (1991) FHB-1 equations - namely the use of indirect measures for the key variables. Specifically, Ruff et al. (1991) used patient-recalled weight as the measure for body mass. While recall information may be useful in some contexts (Olivarius et al., 1997), self-reported weights are notoriously inaccurate, particularly in women (Perry et al., 1995; Bayomi and Tate, 2008). The fact that female mass was incorrectly estimated more often than males in the current sample supports this as a possibility. Ruff et al. (1991) also measured femoral head breadth indirectly from conventional radiographs, compensating for variations in magnification caused by differences in the distance from the radiographic plate, with a single correction factor (19%). While the actual variation between individuals may not have been large, the inability to measure each element directly introduces an additional source of error to the method. Combined with the subjective assessment of weight, this could explain why the equations did not perform well, particularly in females.

The results in relation to the female sub-sample as a whole are more difficult to explain. As noted earlier, none of the FHB equations resulted in mean estimates that met the criteria for acceptance in this group, even when sex-specific equations were used. Apart from the biases discussed above in relation to FHB-1, this initially suggested that the test group females were too different from the females in the reference samples and the equations failed because they were extrapolating beyond their range (Konigsberg et al., 1998). However, the actual samples are not consistent with this explanation. The reference sample for McHenry's (1992) equation had a FHB mean of 41.5 mm and a range of 33-47.5 mm, while Grine et al.'s (1995) reference sample ranged from 38.4-50.5 mm (Ruff, 2010). Our female sample had a FHB mean of 45.5

mm and a range of 39.8-55.5 mm. Thus, it was more similar to Grine et al.'s (1995) reference sample and "fit" the appropriate range for their equation better than McHenry's (1992). As result, Grine et al.'s (1995) FHB-3 equation should have estimated mass better than McHenry's (1992) FHB-2 equation. However, this was not the case and FHB-3 returned higher |PPE|s and estimated fewer individuals within +/-20% of known mass than FHB-2. This indicates that a "match" between the reference sample and target specimen does not ensure a reliable estimate.

Despite this, a reference-target sample "mismatch" does not fully explain why all the FHB equations estimated females relatively poorly. One possibility relates to the level of adipose tissue in females. In addition to carrying more fat and less muscle mass than men, women carry their mass differently and are more prone to fluctuations in weight than men (Shen et al., 2004; Power and Schulkin, 2008). As a result, it is possible that the failure of the postcranial equations to estimate mass as well in females as they did in males has to do with a "looser" relationship between femoral head morphology and body mass in the former. Correlation coefficients in the present sample support this: the relationship between FHB and body mass in females (r=0.15) was considerably lower than that for males (r=0.42). However, on these grounds, the sex-specific FHB-1a (Ruff et al., 1991) and FHB-4a (Ruff et al., 2012) equations should have estimated mass better than the combined-sex equations of McHenry (1992) and Grine et al. (1995) because they were developed from exclusive female reference groups. This was not the case. Thus, further research is needed to determine the cause of the differential results for males and females.

As noted earlier, Lorkiewicz-Muszyńska et al. (2013) also found discrepancies in body mass estimates when using a known-mass sample. Although Lorkiewicz-Muszyńska et al. (2013) did not provide the details of their sample, it was also reasonably large (n=120), presumably European (Polish medical sample), and similar to our sample in age range (20-88 yrs versus our 18-90) and BMI distribution (53% vs our 46% within the "normal range"). Consequently, the pattern of mean differences should have been similar between the two studies. Indeed, in both studies, Grine et al.'s (1995) combined-sex equation estimated mass better than the other equations: females were generally estimated more poorly than males, and the STBIB equations did not return lower mean differences than the FHB equations. However, this means that Lorkiewicz-

Muszyńska et al.'s (2013) results also did not conform to expectations of performance for many of the equations. Consequently, assumptions relating to how existing postcranial equations perform must be re-examined.

Two additional factors suggest that the equations may be even less accurate than what has been described here. First, although many of the equations met the acceptance criteria, the standards were extremely lenient, particularly for intraspecies regressions. As noted earlier, scaling differences between groups suggest that intraspecies regressions will be more accurate than interspecies regressions (Smith, 2002). As a result, a more appropriate criterion for acceptance in an intraspecies analysis would expect the majority of individuals to fall within 10-15% of their known mass (Ruff et al., 2005). However, if we apply this to the current sample and require 50% of the individuals to be estimated within +/-15% of their known mass, the number of "acceptable" equations falls significantly (Appendix Table 11). With a +/-10% criterion, none of the equations would be considered reliable. In fact, no equation estimates more than 41% of the sample within +/-10% of the known mass, a figure that suggests the body mass estimates calculated from these equations should only be considered loose approximations.

The second point relates specifically to the use of the morphometric equations. Although they estimated mass within acceptable limits and performed marginally better than the mechanical equations, it must be emphasized that our results were obtained using documented stature – a condition that is rarely available in skeletal specimens, even modern ones. Importantly, this suggests that existing estimates of body mass that use the morphometric/STBIB method with stature estimated from some other skeletal feature are likely to be even less reliable because the error is compounded. Thus, our results do not support the claim that the STBIB method should necessarily be preferred over the FHB methods, even when bi-iliac breadth and stature can be reliably measured (Auerbach and Ruff, 2004).

The implications of our results for biological anthropology are significant. While most of the equations performed adequately on our study sample, they did not estimate mass well in the female subsample. In addition, the results were achieved under ideal conditions: closely matched test and reference groups, directly measured and

associated variables, and lenient acceptance criteria. For fossil species, whose skeletal elements may require significant reconstruction or approximation and whose body proportions, muscle mass, and sexual dimorphism may differ markedly from modern groups (Churchill et al., 2012), errors in estimation are likely to be much larger than those obtained here. Thus, the existing equations may not be appropriate for the wide range of species they are often applied to (Ruff, 2010) and concerns about making inferences about the physiology and behaviour of fossil hominins from these estimates are still relevant (Smith, 1996). Similar caveats apply to body mass estimates in archaeological and modern samples. Even in forensic contexts, complete elements are rare (Pokines et al., 2013) and in both cases variations in body proportion and muscle mass between groups make it difficult to identify the "appropriate" equation for the target specimen (Ruff et al., 2012). Even with a suitable reference group, fluctuations in individual mass in response to dietary changes, pregnancy, age etc. mean that estimating mass on an individual level will be associated with even greater error.

3.5. Conclusion

Our results suggest that existing equations for estimating body mass from the postcranium need to be used more carefully that they typically have been. Most of the equations met the criteria for acceptance in the test sample, but the limits for acceptance were set very low and reliability dropped significantly when more realistic criteria were used. In addition, not all the equations performed equally well, or within expectations given the sample characteristics. Several assumptions regarding the use of the equations were not fully supported. Specifically, the morphometric/STBIB equations are not more reliable than the mechanical/FHB equations, even when stature and bi-iliac breadth are measured directly. "Matched-target" equations do not consistently estimate mass better than equations designed for broader application. Newer equations for estimating mass from stature-bi-iliac breadth are not generally better, or necessarily more appropriate for high latitude groups, than earlier equations. Although not consistent across all test groups, averaging the results of multiple FHB equations has a neutral or slightly positive effect on estimation accuracy. Sex-specific equations are not necessarily more accurate that equations derived for combined-sexes, but averaging the results of sex-specific equations does not significantly reduce estimation accuracy. Lastly, given

the lenient acceptance criteria employed in the present study, the estimation accuracy for all the equations is likely to be even lower than that achieved here. Consequently, existing body mass estimates derived from these equations must be viewed cautiously.

Our results suggest that current postcranial body mass estimation methods need to be evaluated and applied more critically than has been the practice in biological anthropology to date. In order to do this, the issues that have been identified here must be resolved using large samples of individuals with matched biological information and skeletal data.

Chapter 4.

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

Authors: Marina Elliott, Helen Kurki, Darlene A. Weston and Mark Collard

Submitted: Archaeological and Anthropological Sciences (December 2014)

Statement of Contributions of Joint Authors

Elliott, Marina C. (Candidate): Primary data collection, designing and establishing research methodology, data analysis and interpretation, writing and compilation of manuscript, preparation of figures and tables

Collard, Mark (Senior Supervisor): Supervising and assisting with research design, editing and co-author of manuscript

Kurki, Helen (committee member): assisting with research design, editing and co-author of manuscript

Weston, Darlene (committee member): assisting with research design, editing and co-author of manuscript

This Chapter is an exact copy of the journal paper referred to above

4.1. Introduction

Body mass estimates are crucial for understanding the evolutionary history of humans and other hominins (Damuth and McFadden, 1990; Delson et al., 2000). They are required to understand the adaptive strategies of individual hominin species, and to accurately compare features across fossil groups (Frayer, 1984; Wood and Collard, 1999; Delson et al., 2000; DeSilva, 2011). Body mass estimates are also necessary for interpreting the characteristics of individuals and groups from more recent periods, particularly in terms of growth and development, and health (Steckel and Rose, 2002; Cohen and Crane-Kramer, 2007). Lastly, as a conspicuous individualizing feature and a potential influence on taphonomic processes, body mass estimation is increasingly being used in forensic anthropological research (Suskewicz, 2004; Rainwater et al., 2007; Agostini and Ross, 2011; Byard, 2012).

Over the last quarter of a century, regression analysis has been the main method used to estimate body mass from skeletal remains. This approach regresses a skeletal variable on body mass in samples of extant taxa to generate an equation with which to predict mass in an unknown sample. The equations take the form Y = a + bX, where “Y” is the estimated mass, “X” is the skeletal measurement, “a’’ is the intercept of the regression line, and “b” is its slope. The body masses of unknown specimens are estimated by inserting the target individuals’ values for the skeletal variable into the equation.

Numerous equations for estimating body mass can be found in the literature (Ruff et al., 1991; McHenry, 1992; Aiello and Wood, 1994; Ruff, 1994; Grine et al. 1995; Kappelman 1996; Ruff et al. 1997, 2005; Auerbach and Ruff, 2004, Spocter and Manger, 2007). Specifically, three sets of equations have been derived for cranial variables (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007). Designed to address the paucity of associated postcranial remains in the hominin fossil record and the difficulty of associating such material to a given species, these equations have been used to generate body mass estimates for a variety of fossil hominin and primate specimens (Aiello and Wood, 1994; Kappelman, 1996; Wood and Collard, 1999; Kordos

and Begun, 2001; Rightmire, 2004; Spocter and Manger, 2007; Churchill et al., 2012; Wu and Athreya, 2013). More commonly however, equations are based on postcranial material. The majority of these equations employ measurements of the femur on the grounds that the lower limbs support the weight of the head, torso, and upper limbs and can therefore be expected to directly reflect body mass (Ruff et al., 1991). However, because other factors like activity can also affect femoral morphology (e.g. Lieberman et al., 2004), and because mechanical loading may not be the same in past and present groups (Ruff, 1994), equations based on overall body shape have also been developed (Ruff, 1994; Ruff, 2000; Ruff et al., 2005). These equations employ skeletal variables that correspond to measures of body breadth (bi-iliac breadth) and stature (femoral length), and have been argued to be as good as the femur-based equations (Ruff et al., 1997; Auerbach and Ruff, 2004). These equations have been used widely in palaeoanthropological, bioarchaeological, and forensic contexts (e.g. Ruff and Walker, 1993; Arsuaga et al., 1999; Trinkaus and Jelinek, 1997; Kurki et al., 2010; Myszka et al., 2012).

Despite their widespread use, currently available equations may not be as reliable as they are assumed to be. Evidence for this comes from several recent studies that obtained poor, and/or inconsistent results when testing the equations against populations of known mass (Lorkiewicz-Muszyńska et al., 2013; Elliott et al., 2014; Elliott et al., submitted). In exploring these discrepancies, two issues became apparent as potential sources of error. First, the reference samples used to derive the current equations are problematic. Specifically, sample sizes are often extremely small and the analyses frequently employ indirect measures of key variables and/or unassociated mean data for the skeletal features and body mass (McHenry, 1992; Aiello and Wood, 1994; Kappelman, 1995; Ruff, 2000a; Ruff et al., 2005; Spocter and Manger, 2007). As a result, the true nature of the relationship between the skeletal variables and body size has not been not adequately captured. In addition, validation of the equations has been limited by the paucity of suitable test material: test samples are also often small (e.g. Ruff et al., 1991) and in the absence of known-mass groups, predictive competence is based solely on statistical grounds using correlation coefficients or standard errors of the estimate (e.g. Aiello and Wood, 1994). Other validation efforts have been based on the similarity of results between the morphometric and mechanical methods (Ruff et al, 1997; Auerbach and Ruff, 2004). However, as the samples still include individuals of

unknown mass, these tests serve only to demonstrate the congruence between methods rather than their accuracy. Consequently, the reliability and precision of the current equations for estimating body mass remain uncertain.

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

4.2. Materials and Methods

4.2.1. Sample

This study used archived CT data from 128 male and 125 female deceased modern humans (range = 18-90 years, male mean = 48.1 yrs, female mean = 51.2 yrs). CT scans were conducted at the Institute of Forensic Medicine (IFM) at the University of Zurich, Switzerland as part of routine forensic analyses (Thali et al., 2003) and the data were accessed through the IFM's secure server in accordance with their protocols. Query searches, record reviews and visual inspection of the CT scans were used to select the sample. Exclusion criteria included individuals with skeletal abnormalities, trauma or prosthetics in the anatomical regions of interest, and any individual who was processed more than three days after death. Sex, age (yrs), body mass (kg) and stature (cm) were recorded for each individual. The full sample of 253 individuals was divided into two groups (Tables 4.1 and 4.2): a "calibration" (n=203) and a randomly chosen "test" (n=50) sample, of roughly equal numbers of males and females each. The calibration sample was used to derive the regression equations, while the test sample was used to evaluate their accuracy and precision. Although other methods have been suggested (Smith, 2002), such an external validation continues to be the most unbiased and rigorous means of testing predictive equations (Harrell et al 1996; Porter, 1999; Giancristofaro and Salmaso, 2007).

Table 4.1. Summary statistics for calibration sample (n=203)

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

Table 4.2. Summary statistics for test sample (n=50)

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

4.2.2. Imaging and 3D reconstruction protocols

CT scans were conducted at the IFM using a 128-slice, Siemens SOMATOM® Definition Flash, Dual-source CT scanner (Siemens Healthcare; Forchheim, Germany). Scans were accessed from the IFM archive and anatomical regions volume rendered using OsiriX imaging software (64-bit extension, http://www.osirix-viewer.com) (Figure 1). Skeletal elements were oriented in consistent planes and measured on the right side to the nearest 0.1 mm using OsiriX tools. The accuracy of volume rendering skeletal models from CT has been demonstrated by several studies (Decker et al., 2011; Kim et al., 2012; Smyth et al., 2012). It was verified during this project by measuring, scanning, virtually reconstructing, and then virtually re-measuring an archaeological skull from the IFM's collection (Elliott et al., 2014). Measurement differences between the physical and virtual measurements in the latter test were less than 3%.

Figure 4.1. Example of a three-dimensionally reconstructed skeletal element, volume-rendered from CT data by OsiriX

4.2.3. Skeletal variables

Twelve cranial and 24 postcranial variables were used for this study (Tables 4.3, 4.4, Fig 4.2). The cranial variables included six linear measurements selected on the basis of their performance in previous studies (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007). All of these studies identified orbital and foramen magnum areas as useful predictors of body mass in hominoids. However, as the variables were not measured in the same way, three calculations each for orbital and foramen magnum area were included here: simple breadth x height (area = b x h), in accordance with Aiello and Wood (1994) and Spocter and Manger (2007); area as an ellipse (area = (π/4) x b x h), following Spocter and Manger (2007); and a CAD-assisted method that traced the perimeter of the feature from a two-dimensional image, using a procedure similar to that of Kappelman (1996).

Table 4.3. Cranial variables used in this study

	Cranial Variables	Description	Reference
1	BORB	Orbital breadth: distance between maxillofrontale and ectoconchion – in mm	Aiello & Wood 1994
2	HORB	Orbital height: distance between superior and inferior orbital margins, taken at a right angle to BORB – in mm	Aiello & Wood 1994
3	BIOR	Biorbital breadth: distance between two ectoconchion – in mm	Aiello & Wood 1994
4	BPOR	Biporionic breadth: distance from porion to porion – in mm	Aiello & Wood 1994
5	LFM	Foramen magnum length: distance between basion and opisthion – in mm	Aiello & Wood 1994
6	BFM	Foramen magnum breadth: distance in the coronal plane between the inner margins of the foramen magnum – in mm	Aiello & Wood 1994

	Cranial Variables	Description	Reference
7	ORBA1	Orbital area (b x h): product of breadth x height – in mm^2	Aiello & Wood 1994
8	ORBA2	Orbital area (ellipse): calculated from breadth x height as an ellipse – in mm^2	Spocter & Manger 2007
9	ORBA3	Orbital area (CAD): calculated from perimeter margin using area function of ImageJ – in mm^2	Kappelman 1996
10	FMA1	Foramen magnum area (b x h): product of breadth x height – in mm^2	Aiello & Wood 1994
11	FMA2	Foramen magnum area (ellipse): calculated from breadth x height as an ellipse – in mm^2	Spocter & Manger 2007
12	FMA3	Foramen magnum area (CAD): calculated from perimeter margin using area function of ImageJ – in mm^2	This study

Table 4.4. Postcranial variables used in this study

	Postcranial variables	Description	Reference
13	FHB	Femoral head breadth: superior-inferior breadth perpendicular to the cervical axis – in mm	Ruff et al. 1991
14	BIB	Bi-iliac breadth: maximum pelvic breadth taken across the iliac crests, taken as "Living BIB" – in mm	Ruff, 1991, 1994
15	FNB	Femoral neck breadth: minimum superior-inferior breadth at the point of deepest concavity of the superior surface – in mm	Ruff et al. 1991
16	FLM	Max femoral length: length of femur from the most superior point on the head to the most inferior point on the distal condyles - in mm	Buikstra & Ubelaker 1994
17	MLSB80	Medio-lateral shaft breadth at 80% of total femur length, as measured from the distal end - in mm	Ruff and Hayes 1983
18	MLCB80	Sum of medial and lateral cortical breadths at 80% of total femur length, as measured from the distal end - in mm	Ruff and Hayes 1983
19	CA80	Index of cortical area at 80%: calculated from medio-lateral shaft (D) and cortical (d) breadth using $\pi/4(D^2-d^2)$ – in mm^2	Ruff et al. 1991
20	I80	Index of second moment of area at 80%: calculated from shaft (D) and cortical breadth (d) using $\pi/64(D^4-d^4)$ – in mm^4	Ruff et al. 1991
21	MLSB65	Medio-lateral shaft breadth at 65% of total femur length, as measured from the distal end – in mm	Ruff and Hayes 1983
22	MLCB65	Sum of medial and lateral cortical breadths at 65% of total femur length, as measured from the distal end – in mm	Ruff and Hayes 1983
23	CA65	Index of cortical area at 65%: calculated from shaft (D) and cortical (d) breadth using $\pi/4(D^2-d^2)$ – in mm^2	Ruff et al. 1991

	Postcranial variables	Description	Reference
24	I65	Index of second moment of area at 65%: calculated from shaft (D) and cortical breadth (d) using $\pi/64(D^4-d^4)$ – in mm^4	Ruff et al. 1991
25	MLSB50	Medio-lateral shaft breadth at 50% of total femur length, as measured from the distal end – in mm	Ruff and Hayes 1983
26	MLCB50	Sum of medial and lateral cortical breadths at 50% of total femur length, as measured from the distal end – in mm	Ruff and Hayes 1983
27	CA50	Index of cortical area at 50%: calculated from shaft (D) and cortical (d) breadth using $\pi/4(D^2-d^2)$ – in mm^2	Ruff et al. 1991
28	I50	Index of second moment of area at 50%: calculated from shaft (D) and cortical breadth (d) using $\pi/64(D^4-d^4)$ – in mm^4	Ruff et al. 1991
29	MLSB35	Medio-lateral shaft breadth at 35% of total femur length, as measured from the distal end – in mm	Ruff and Hayes 1983
30	MLCB35	Sum of medial and lateral cortical breadths at 35% of total femur length, as measured from the distal end – in mm	Ruff and Hayes 1983
31	CA35	Index of cortical area at 35%: calculated from shaft (D) and cortical (d) breadth using $\pi/4(D^2-d^2)$ – in mm^2	Ruff et al. 1991
32	I35	Index of second moment of area at 35%: calculated from shaft (D) and cortical breadth (d) using $\pi/64(D^4-d^4)$ – in mm^4	Ruff et al. 1991
33	MLSB20	Medio-lateral shaft breadth at 20% of total femur length, as measured from the distal end – in mm	Ruff and Hayes 1983
34	MLCB20	Sum of medial and lateral cortical breadths at 20% of total femur length, as measured from the distal end – in mm	Ruff and Hayes 1983
35	CA20	Index of cortical area at 20%: calculated from shaft (D) and cortical (d) breadth using $\pi/4(D^2-d^2)$ – in mm^2	Ruff et al. 1991
36	I20	Index of second moment of area at 20%: calculated from shaft (D) and cortical breadth (d) using $\pi/64(D^4-d^4)$ – in mm^4	Ruff et al. 1991

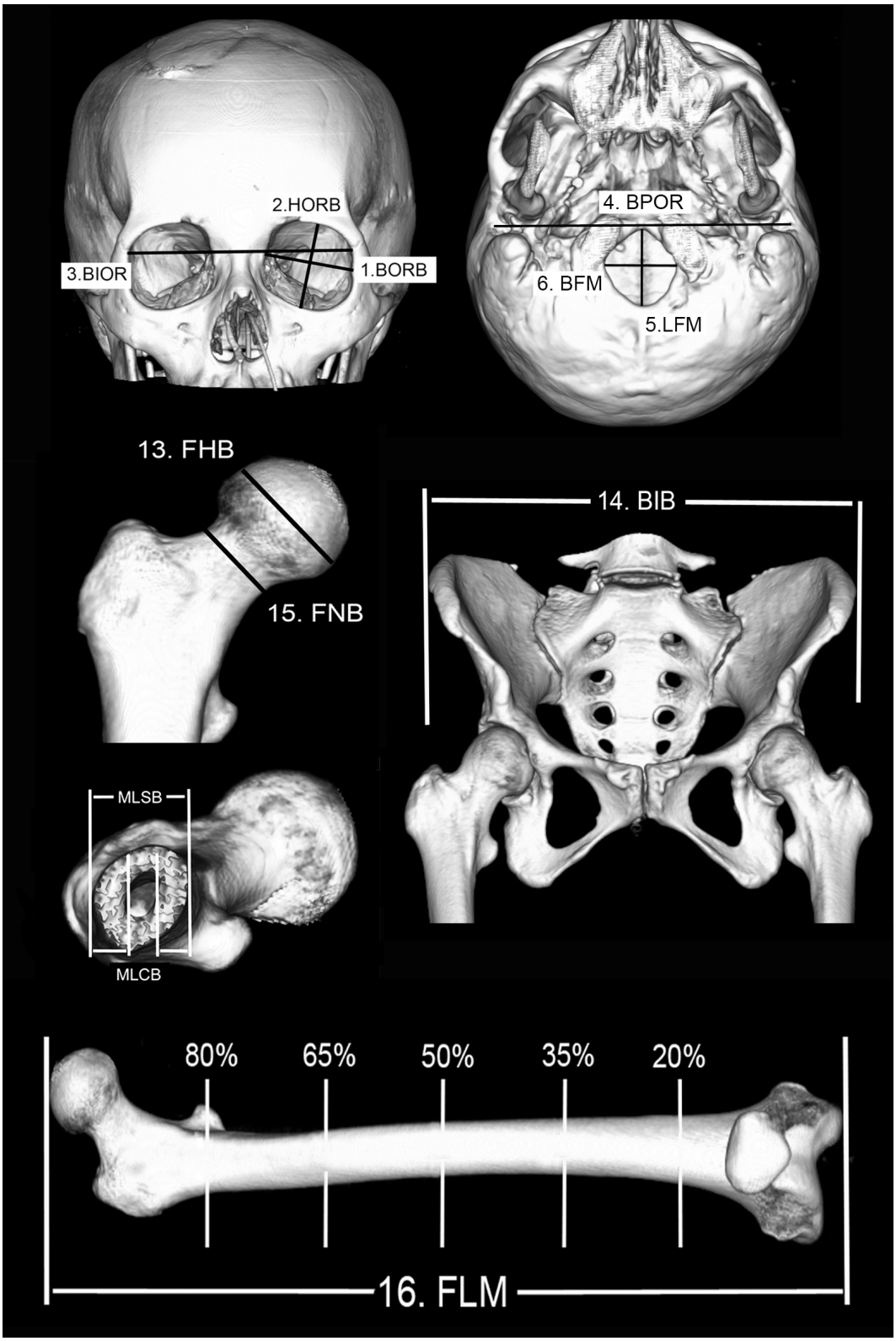


Figure 4.2. Cranial and post-cranial measurement variables and cross-sectional locations used in this study

The postcranial variables included the two most commonly used measurements for predicting body mass from skeletal material: femoral head breadth (FHB) and bi-iliac breadth (BIB). FHB has been used in four sets of published regression equations (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995; Ruff et al., 2012) and is associated with the "mechanical" estimation approach (Ruff, 2002; Auerbach and Ruff, 2004). BIB is used in conjunction with stature (STAT), in second set of postcranial equations known as the "morphometric" approach (Ruff, 1994; Ruff et al., 1997; 2005). An additional 12 measurements of the femur were recorded in order to examine other assertions that have been made in the literature. Specifically, femoral neck breadth (FNB) was measured in light of Ruff's (1991) suggestion that femoral neck size will exhibit a pattern of correlation with body mass between that of the head and shaft due to its intermediate location. Following Ruff and Hayes (1983) and Ruff (1991; 2000b), Maximum femoral length (FLM) and measurements of medio-lateral shaft breadth (MLSB) and cortical breadth (MLCB), taken at multiple locations along the femur, were recorded to explore how diaphyseal cross-sectional dimensions relate to weight. Lastly, shaft and cortical breadths were used to calculate 10 indices that correspond to cortical area (CA) and second moment of area (I) (Ruff et al., 1991). CA and I provide information about the cross-sectional geometry and strength of the femur (Ruff and Hayes, 1983) and are calculated as: $CA=\pi/4(D2-d2)$ and $I=\pi/64(D4-d4)$, respectively. Although the femur is obviously not cylindrical, and the indices are therefore only approximations of the 'true' cross-sectional value of the element, these calculations were used here for direct comparison with Ruff and Hayes (1983) and Ruff et al. (1991). Tables 4.5-4.8 provide the summary data for the 36 variables in both samples.

Table 4.5. Summary data for cranial variables, calibration sample (n=203)

	Female (n=100)		Male (n=103)		Combined-sex (n=203)	
Variable	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
BORB	35.2 (1.6)	31.9-39.7	37.2 (1.9)	33.7-42.3	36.2 (2.0)	31.9-42.3
HORB	34.3 (2.2)	28.3-39.9	34.3 (2.3)	27.5-41.4	34.3 (2.3)	27.5-41.4
BIOR	93.7 (4.0)	84.7-105.9	98.4 (4.1)	88.3-107.1	96.0 (4.7)	84.7-107.1
BPOR	112.3 (5.1)	99.5-123.3	119.9 (5.1)	109.6-132.3	116.1 (6.4)	99.5-132.3
LFM	33.6 (2.2)	29.4-41.2	35.1 (2.6)	30.0-42.1	34.4 (2.5)	29.4-42.1
BFM	28.5 (92.3)	22.9-34.1	29.7 (2.5)	22.1-37.3	29.1 (2.5)	22.1-37.3
ORBA1	1207.7 (104.0)	911.3-1501.5	1275.5 (107.6)	1035.0-1531.9	1242.1 (110.9)	911.3-1531.9
ORBA2	948.6 (81.7)	715.7-1179.3	1001.8 (84.5)	812.9-1203.1	975.6 (87.1)	715.7-1203.1
ORBA3	969.6 (81.4)	766.2-1205.7	1024.3 (81.7)	831.7-1267.9	997.2 (85.8)	766.2-1267.9
FMA1	961.2 (125.6)	692.3-1404.9	1045.3 (150.5)	758.0-1507.2	1003.9 (144.7)	692.3-1507.2
FMA2	754.9 (98.7)	543.7-1103.4	821.0 (118.2)	595.4-1183.7	788.4 (113.7)	543.7-1183.7
FMA3	679.6 (90.3)	490.5-1008.8	745.7 (102.3)	548.1-1059.1	713.0 (101.8)	490.5-1059.1

Table 4.6. Summary data for postcranial variables, calibration sample (n=203)

	Female (n=100)		Male (n=103)		Combined-sex (n=203)	
Variable	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
BIB	279.2 (19.2)	223.1-337.1	284.8 (15.5)	246.3-324.2	282.0 (17.6)	223.1-337.1
FLM	445.0 (26.1)	390.1-530.3	474.1 (25.6)	412.3-534.6	459.8 (29.6)	390.1-534.6
FHB	45.6 (2.4)	39.8-55.5	51.1 (2.6)	43.2-57.5	48.4 (3.7)	39.8-57.5
FNB	31.9 (2.3)	26.7-38.6	36.6 (2.4)	30.1-43.2	34.3 (3.3)	26.7-43.2
MLSB80	33.8 (2.2)	27.7-41.2	37.3 (2.6)	30.9-43.2	35.6 (3.0)	27.7-43.2
MLCB80	16.4 (2.5)	10.6-23.9	17.5 (3.1)	8.4-26.1	16.9 (2.9)	8.4-26.1
CA80	686.0 (104.7)	404-1187.9	853.8 (127.3)	593.8-1179.6	771.1 (143.6)	404.1-1187.9
I80	$6.2x10^4$ ($1.7x10^4$)	$2.6x10^4$-$1.4x10^5$	$9.3x10^4$ ($2.6x10^4$)	$4.4x10^4$-$1.6x10^5$	$7.8x10^4$ ($2.7x10^4$)	$2.6x10^4$-$1.6x10^5$
MLSB65	29.0 (2.4)	24.5-35.4	32.2 (2.2)	26.1-38.4	30.6 (2.8)	24.5-38.4
MLCB65	8.2 (2.0)	4.8-13.8	9.2 (2.1)	5.0-19.9	8.7 (2.1)	4.8-19.9
CA65	609.2 (109.3)	424.9-920.0	749.2 (107.7)	482.2-1079.6	680.2 (129.0)	424.9-1079.6
I65	$3.6x10^4$ ($1.3x10^4$)	1.8x4-$7.6x10^4$	$5.4x10^4$ ($1.5x10^4$)	$2.3x10^4$-$1.1x10^5$	$4.5x10^4$ ($1.6x10^4$)	$1.8x10^4$-$1.1x10^5$
MLSB50	28.6 (2.1)	24.0-35.1	31.6 (2.0)	25.6-37.6	30.1 (2.5)	24.0-37.6
MLCB50	8.8 (1.8)	3.5-12.6	9.7 (1.9)	5.7-13.9	9.2 (1.9)	3.5-13.9
CA50	585.2 (95.7)	390.2-860.1	709.5 (93.0)	468.2-1004.7	648.3 (112.9)	390.2-1004.7
I50	$3.4x10^4$ ($1.1x10^4$)	$1.6x10^4$-$7.4x10^4$	$4.9x10^4$ ($1.3x10^4$)	$2.1x10^4$-$9.7x10^4$	$4.2x10^4$ ($1.4x10^4$)	$1.6x10^4$-$9.7x10^4$
MLSB35	30.8 (2.3)	25.7-36.2	34.1 (2.2)	30.1-40.7	32.5 (2.8)	25.7-40.7

	Female (n=100)		Male (n=103)		Combined-sex (n=203)	
Variable	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
MLCB35	14.7 (2.7)	7.1-21.6	15.9 (2.8)	9.2-23.5	15.3 (2.8)	7.1-23.5
CA35	575.0 (84.0)	380.0-787.9	713.8 (90.7)	540.6-1097.4	645.4 (111.6)	380.0-1099.4
I35	$4.3x10^4$ ($1.3x10^4$)	$2.1x10^4$-$7.9x10^4$	$6.4x10^4$ ($1.7x10^4$)	$3.9x10^4$-$1.3x10^5$	$5.4x10^4$ ($1.8x10^4$)	$2.1x10^4$-$1.3x10^5$
MLSB20	39.8 (3.2)	32.4-49.1	43.1 (3.5)	36.2-51.9	41.5 (3.7)	32.4-51.9
MLCB20	26.7 (4.1)	13.7-39.4	28.0 (4.6)	13.8-39.5	27.4 (4.4)	13.7-39.5
CA20	677.4 (124.8)	394.5-1092.6	836.4 (154.4)	529.9-1499.8	758.1 (161.3)	394.5-1499.8
I20	$9.9x10^4$ ($2.9x10^4$)	$4.4x10^4$-$1.8x10^5$	$1.4x10^4$ ($4.4x10^4$)	$7.3x10^4$-$3.3x10^5$	$1.2x10^5$ ($4.3x10^4$)	$4.4x10^4$-$3.3x10^5$

Table 4.7. Summary data for cranial variables, test sample (n=50)

	Female (n=25)		Male (n=25)		Combined-sex (n=50)	
Variable	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
BORB	35.4 (1.8)	32.0-39.4	36.8 (1.3)	34.8-39.3	36.1 (1.7)	32.0-39.4
HORB	33.5 (2.1)	28.6-36.8	34.5 (1.8)	30.6-37.5	34.0 (2.0)	28.6-37.5
BIOR	93.8 (3.2)	87.4-100.7	97.8 (3.0)	92.7-102.4	95.8 (3.7)	87.4-102.4
BPOR	111.9 (4.6)	105.7-120.5	118.4 (4.7)	110.7-128.9	115.1 (5.6)	105.7-128.9
LFM	32.6 (1.8)	29.6-36.5	34.8 (2.8)	30.5-41.1	33.7 (2.6)	29.6-41.1
BFM	27.9 (1.6)	24.9-31.1	29.8 (2.4)	26.1-34.8	28.8 (2.2)	24.9-34.8
ORBA1	1189.1 (105.4)	915.2-1335.3	1271.4 (85.0)	1092.4-1409.1	1230.2 (103.5)	915.2-1409.1
ORBA2	933.9 (82.8)	718.8-1048.8	998.5 (66.8)	858.0-1106.7	966.2 (81.3)	718.8-1106.7
ORBA3	957.8 (87.1)	732.8-1076.2	1008.4 (78.2)	837.4-1146.3	982.1 (86.0)	732.8-1146.3
FMA1	910.2 (75.4)	744.5-1042.6	1039.5 (142.7)	802.2-1430.3	974.9 (130.5)	744.5-1430.3
FMA2	714.9 (59.2)	584.7-818.8	816.4 (112.1)	630.0-1123.3	765.6 (102.5)	584.7-1123.3
FMA3	638.6 (62.1)	504.7-739.3	727.7 (103.9)	583.2-982.1	681.3 (95.1)	504.7-982.1

Table 4.8. Summary data for postcranial variables, test sample (n=50)

	Female (n=25)		Male (n=25)		Combined-sex (n=50)	
Variable	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
BIB	272.9 (14.3)	251.6-302.5	280.0 (21.1)	211.2-313.2	276.5 (18.2)	211.2-313.2
FLM	433.3 (20.8)	384.7-494.6	475.8 (27.1)	431.1-549.1	454.6 (32.1)	384.7-549.1
FHB	45.0 (1.8)	41.5-48.1	50.3 (3.5)	41.5-54.9	47.6 (3.9)	41.5-54.9
FNB	31.6 (2.1)	26.7-35.8	36.0 (3.2)	28.3-43.8	33.8 (3.5)	26.7-43.8
MLSB80	33.3 (2.2)	29.1-38.3	36.2 (3.3)	26.7-42.0	34.8 (3.1)	26.7-42.0
MLCB80	15.5 (1.9)	12.2-18.9	17.6 (3.1)	11.7-26.4	16.5 (2.8)	11.7-26.4
CA80	684.8 (96.5)	504.5-973.0	787.4 (175.4)	452.4-1122.4	736.1 (149.7)	452.4-1122.4
I80	$9.4x10^5$ ($2.5x10^5$)	$5.3x10^5$-$1.6x10^6$	$1.3x10^6$ ($4.6x10^5$)	$3.8x10^5$-$2.4x10^6$	$1.1x10^6$ ($4.1x10^5$)	$3.8x10^5$-$2.4x10^6$
MLSB65	28.8 (2.3)	25.2-35.6	30.9 (2.9)	23.0-36.7	29.8 (2.8)	23.0-36.7
MLCB65	7.6 (1.5)	4.1-10.6	8.6 (2.1)	3.2-12.5	8.1 (1.9)	3.2-12.5
CA65	606.7 (109.3)	464.5-975.0	694.6 (138.7)	389.1-987.0	650.7 (131.3)	389.1-987.0
I65	$5.6x10^5$ ($2.0x10^5$)	$3.2x10^5$-$1.3x10^6$	$7.5x10^5$ ($2.7x10^5$)	$2.2x10^5$-$1.4x10^6$	$6.5x10^5$ ($2.6x10^5$)	$2.2x10^5$-$1.4x10^6$
MLSB50	28.3 (2.2)	24.2-34.4	30.7 (2.8)	25.1-35.8	29.5 (2.7)	24.2-35.8
MLCB50	8.4 (1.9)	3.6-11.7	9.2 (1.5)	6.3-12.9	8.8 (1.8)	3.6-12.9
CA50	576.0 (96.7)	427.8-889.8	677.0 (134.5)	453.0-940.1	626.5 (126.7)	427.8-940.1
I50	$5.2x10^5$ ($1.7x10^5$)	$2.7x10^5$-$1.1x10^6$	$7.2x10^5$ ($2.7x10^5$)	$31.x10^5$-$1.3x10^6$	$6.2x10^5$ ($2.4x10^5$)	$2.7x10^5$-$1.3x10^6$
MLSB35	30.5 (2.4)	25.7-34.5	33.2 (3.0)	27.1-40.5	31.8 (3.0)	25.7-40.5

	Female (n=25)		Male (n=25)		Combined-sex (n=50)	
Variable	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
MLCB35	14.0 (2.8)	8.4-19.8	15.0 (2.7)	11.5-20.4	14.5 (2.8)	8.4-20.4
CA35	574.0 (88.2)	409.4-812.1	689.5 (124.0)	468.0-1055.9	631.8 (121.5)	409.4-1055.9
I35	$6.6x10^5$ ($1.9x10^5$)	$3.3x10^5$-$1.1x10^6$	$9.5x10^5$ ($3.5x10^5$)	$4.1x10^5$-$2.0x10^6$	$8.1x10^5$ ($3.2x10^5$)	$3.3x10^5$-$2.0x10^6$
MLSB20	39.1 (3.9)	32.4-45.4	41.9 (4.9)	33.5-52.0	40.5 (4.6)	32.4-52.0
MLCB20	25.8 (3.7)	19.5-34.8	27.2 (4.9)	19.0-37.8	26.5 (4.3)	19.0-37.8
CA20	678.4 (128.7)	464.8-1010.0	799.2 (182.7)	529.3-1234.6	738.8 (167.9)	464.8-1234.6
I20	$1.5x10^6$ ($5.5x10^5$)	$7.3x10^5$-$2.7x10^6$	$2.1x10^6$ ($9.5x10^5$)	$8.6x10^5$-$4.7x10^6$	$1.8x10^6$ ($8.2x10^5$)	$7.3x10^5$-$4.7x10^6$

4.2.4. Analyses

Using the calibration sample (n=203), correlation coefficients were calculated and regression equations derived for each variable for a combined-sex sample, as well as female-only (n=100) and male-only (n=103) subsamples. Regression equations were derived for each group using three methods: least square (LS), reduced major axis (RMA) and major axis (MA) regression because debate continues to surround the most appropriate choice for predicting an unknown quantity using regression. Konigsberg et al. (1998) outline five univariate estimation methods ("inverse calibration", "classic calibration", "major axis regression", "reduced major axis regression" and "ratio estimators") and note that least square regression (LRS) of body size (stature) on bone length (termed "inverse calibration") is the most appropriate when the target specimen can be assumed to come from the same distribution as the calibration sample. Alternatively, they suggest that "classic calibration" – regression of bone length on stature, followed by solving for stature – may be better in situations where extrapolation is expected or uncertain. Despite this, most existing equations for estimating body mass have employed LSR and "inverse calibration" on the grounds that it minimizes the estimation error of the dependent variable (in this case, body mass) (Aiello and Wood, 1994). LSR also has the advantage of having correction factors available to account for

biases inherent in the technique (Smith, 1996). Alternate methods like Reduced Major Axis (RMA) or Major axis (MA), have been argued to accept more uncertainty in the variables and produce better results when extrapolating beyond the range of the original dataset (Ruff et al., 1991, Auerbach and Ruff, 2004). However, body mass estimates continue to be derived from one or other method (Pomeroy and Stock, 2012; Ruff et al., 2012) and consensus on the most appropriate approach has not been reached (Smith, 2009). Consequently, we derived equations with all three methods. To avoid problems associated with non-normal distributions, the data were logarithmically transformed (base 10). Analyses were also conducted on the raw data, but as the results were generally worse, they have not been included here. The standard error of the estimate (SEE) was calculated according to the formula SEE= $\sqrt{\sum}(y-y')2/n-2$ (Hinton, 2004). Two correction factors - the smearing estimate (SE) and ratio estimator (RE) - were also calculated to account for de-transformation biases when converting from logarithmically transformed units into arithmetic units in LS regression (Smith, 1993).

Once the regression equations were derived, individual skeletal measurements from the test sample were entered into the corresponding equation and an estimated body mass calculated in kilograms. For the LS regression equations the estimated mass was corrected by multiplying it by the average of the SE and RE, to obtain a final estimated body mass (EBM) (Smith, 1993). Estimated body masses were then compared to known mass for each individual. Raw differences (RD) were calculated as known-EBM. Percentage differences were calculated as percent error using the formula (PE)= (known-EBM)/ known*100 (Wu et al., 1995). Absolute percentage differences (|PE|) were also calculated to assess the magnitude of the difference between the estimated and known masses (Aiello and Wood, 1994; Ruff et al., 2005). Mean RDs, PEs and |PE|s were calculated for males and females as well as the whole test sample. Differences were plotted and Wilcoxon signed-rank tests used to establish their significance. Lastly, in keeping with previous studies (e.g. Dagosto and Terranova, 1992; Aiello and Wood, 1994), we also calculated the percentage of individuals whose EBM fell within +/-20% of their known mass. All analyses were conducted in “R” (R Development Core Team, 2010).

4.2.5. Evaluation criteria

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

4.3. Results

4.3.1. Regression equations

Combined-sex test sample: Tables 4.9 and 4.10 provide the LSR, MA and RMA regression statistics derived from the combined-sex calibration sample. Correlation coefficients (r) for the cranial variables range from 0.06 (p=0.36) for orbital height (HORB) to 0.33 (p=0.00) for biporionic breadth (BPOR). For the postcranial variables, correlation coefficients range from -0.03 (p=0.63) for mediolateral cortical breadth at

80% of femoral shaft length (MLCB80) to 0.59 (p=0.00) for the index of cortical area at 50% of femoral shaft length (CA50). SEEs for the cranial variables are consistent at 0.11, while SEEs vary slightly for the postcranial variables (0.09-0.11).

Table 4.9. Least square (LS), Major Axis (MA) and Reduced Major Axis (RMA) regression equations for cranial variables. Combined-sex calibration sample (n=203). All data log10 transformed. See= Standard error of estimate, CF=Correction factor (mean of SE and RE)

			LSR regression				MA regression		RMA regression	
Variable	**r**	**p-value**	**Slope**	**Intercept**	**See**	**CF**	**Slope**	**Intercept**	**Slope**	**Intercept**
BORB	0.21	0.003	0.95	0.38	0.11	1.03	21.45	-31.59	4.62	-5.34
HORB*	0.06	0.362	0.24	1.49	0.11	1.03	55.24	-82.89	3.81	-3.98
BIOR	0.30	0.000	1.57	-1.24	0.11	1.03	16.67	-31.18	5.20	-8.43
BPOR	0.33	0.000	1.53	-1.28	0.11	1.03	13.23	-25.44	4.59	-7.61
LFM	0.21	0.012	0.73	0.75	0.11	1.03	15.59	-22.06	3.51	-3.52
BFM*	0.13	0.063	0.38	1.30	0.11	1.03	19.92	-27.27	2.94	-2.43
ORBA1	0.17	0.013	0.49	0.34	0.11	1.03	14.34	-42.47	2.83	-6.89
ORBA2	0.17	0.013	0.49	0.40	0.11	1.03	14.34	-40.96	2.83	-6.59
ORBA3	0.18	0.012	0.52	0.30	0.11	1.04	14.81	-42.51	2.95	-6.97
FMA1	0.19	0.007	0.34	0.84	0.11	1.03	6.86	-18.69	1.81	-3.57
FMA2	0.18	0.007	0.34	0.88	0.11	1.03	6.86	-17.97	1.81	-3.38
FMA3	0.20	0.004	0.37	0.81	0.11	1.03	6.50	-16.65	1.83	-3.34

* regression non-significant at $\propto$=0.05

Table 4.10. Least square (LS), Major Axis (MA) and Reduced Major Axis (RMA) regression equations for postcranial variables. Combined-sex sample (n=203). All data log10 transformed. See= Standard error of estimate, CF=Correction factor (mean of SE and RE)

			LSR regression				MA regression		RMA regression	
Variable	r	p-value	Slope	Intercept	See	CF	Slope	Intercept	Slope	Intercept
BIB	0.36	0.000	1.43	-1.65	0.10	1.03	10.6	-24.1	4.0	-7.96
FLM	0.33	0.000	1.28	-1.54	0.10	1.03	11.3	-28.1	3.9	-8.54
FHB	0.40	0.000	1.33	-0.38	0.10	1.03	7.6	-11.0	3.3	-3.72
FNB	0.43	0.000	1.12	0.16	0.10	1.03	5.3	-6.3	2.6	-2.11
MLSB80	0.41	0.000	1.25	-0.06	0.10	1.03	6.6	-8.4	3.0	-2.78
MLCB80*	-0.03	0.630	- 0.05	1.93	0.11	1.03	-23.3	30.3	-1.5	3.71
CA80	0.52	0.000	0.71	-0.18	0.09	1.03	1.8	-3.3	1.4	-2.11
I80	0.44	0.000	0.33	0.24	0.10	1.03	0.54	-0.8	0.8	-1.8
MLSB65	0.49	0.000	1.35	-0.14	0.10	1.03	5.1	-5.6	2.7	-2.20
MLCB65	0.14	0.039	- 0.15	2.00	0.11	1.03	-1.4	3.1	-1.0	2.82
CA65	0.55	0.000	0.73	- 0.18	0.09	1.03	1.6	-2.7	1.3	-1.91
I65	0.50	0.000	0.34	0.28	0.10	1.03	0.5	-0.4	0.7	-1.30
MLSB50	0.52	0.000	1.58	-0.47	0.09	1.03	5.3	-6.0	3.0	-2.6
MLCB50	0.15	0.038	-0.17	2.02	0.11	1.03	-2.3	4.1	-1.1	2.96
CA50	0.59	0.000	0.85	-0.53	0.09	1.03	1.8	-3.2	1.4	-2.18
I50	0.53	0.000	0.40	0.03	0.09	1.03	0.6	-0.9	0.8	-1.60
MLSB35	0.44	0.000	1.29	-0.08	0.10	1.03	6.1	-7.4	2.9	-2.58
MLCB35*	0.00	0.992	0.00	1.87	0.11	1.03	0.0	-844.9	0.0	0.33
CA35	0.56	0.000	0.82	-0.44	0.09	1.03	1.9	-3.5	1.5	-2.23
I35	0.47	0.000	0.35	0.21	0.97	1.03	0.6	-0.7	0.7	-1.65
MLSB20	0.32	0.000	0.89	0.43	0.10	1.03	7.8	-10.8	2.8	-2.66
MLCB20*	0.04	0.597	0.06	1.79	0.11	1.03	22.3	-30.0	1.5	-0.28
CA20	0.50	0.000	0.62	0.10	0.10	1.03	1.5	-2.5	1.2	-1.68
I20	0.42	0.000	0.31	0.32	0.10	1.03	0.5	-0.7	0.7	-1.84

* regression non-significant at $\propto$=0.05

Female test sample: Tables 4.11 and 4.12 provides the regression equations derived from the female (n=100) calibration sample. Here, the correlation coefficients for the cranial measurements vary from -0.01 for foramen magnum breadth (BFM) to 0. 12 for biporionic breadth (BPOR). However, none of the correlations is significant at $\propto$=0.05. Correlation coefficients for the postcranial measurements vary from -0.11 (p=0.28) for mediolateral cortical breadth at 35% (MLCB35) to 0.59 (p=0.00) for the index of cortical area at 50% of femoral shaft length (CA50). SEEs for the cranial variables are again consistent, this time at 0.12. SEEs for the postcranial variables range from 0.10-0.12.

Table 4.11. Least square (LS), Major Axis (MA) and Reduced Major Axis (RMA) regression equations for cranial variables. Female calibration sample (n=100). All data log10 transformed. See= Standard error of estimate, CF=Correction factor (mean of SE and RE)

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

* regression non-significant at $\propto$=0.05

Table 4.12. Least square (LS), Major Axis (MA) and Reduced Major Axis (RMA) regression equations for postcranial variables. Female calibration sample (n=100). All data log10 transformed. See= Standard error of estimate, CF=Correction factor (mean of SE and RE)

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

* regression non-significant at $\propto$=0.05

Male test sample: In the male sample (Tables 4.13 and 4.14), the correlation coefficients for the cranial measurements vary from 0.05 (p=0.62) for orbital height (HORB), to 0.28 (p=0.00) for biporionic breadth (BPOR). Correlations for the postcranial variables range from -0.02 (p=0.81) for medio-lateral cortical breadth at 35% of femoral shaft (MLCB35), to 0.42 (p=0.00) for the index of cortical area at 50% of femoral shaft length (CA50). SEEs for the cranial variables are consistent at 0.09 and range from 0.08-0.09 for the postcranial variables.

Table 4.13. Least square (LS), Major Axis (MA) and Reduced Major Axis (RMA) regression equations for cranial variables. Male calibration sample (n=103). All data log10 transformed. See= Standard error of estimate, CF=Correction factor (mean of SE and RE)

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

* regression non-significant at $\propto$=0.05

Table 4.14. Least square (LS), Major Axis (MA) and Reduced Major Axis (RMA) regression equations for postcranial variables. Male calibration sample (n=103). All data log10 transformed. See= Standard error of estimate, CF=Correction factor (mean of SE and RE)

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

* regression non-significant at $\propto$=0.05

4.3.2. Prediction accuracy

Tables 4.15-4.20 summarize the directional and absolute differences, as well as the percentage of individuals whose body mass was estimated within +/-20% of their true mass, for each sample using the LSR regression equations. It must be noted here that a number of the regressions in each training group were non-significant at $\propto$=0.05. Normally, this would preclude further use of the resulting equation for predictive purposes. However, for maximum comparability with previous studies and heuristic purposes, all of the regression equations were tested as derived. As the major axis (MA) and Reduced Major Axis (RMA) regression equations produced significantly higher rates of error for most skeletal measurements in all three samples, their results are provided in the Appendix (Tables A12-13) and only the least square (LS) regression results are discussed here.

Combined-sex test sample: In this sample, the cranial variables with the best predictive accuracy based on the absolute percentage errors and number of individuals estimated within +/-20% of known mass, were biporionic breadth (BPOR) and biorbital breadth (BIOR) (Tables 4.15 and 4.16). These two variables resulted in mean absolute percent errors (|PE|) of 17.1 and 18.1% respectively. Both variables estimated 64% of the test sample within +/-20% of the known mass. The worst performing cranial variable was foramen magnum breadth, which produced a |PE| of 19.6% and estimated 56% of the sample within +/-20% of known mass.

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

Table 4.15. Summary of difference between known and estimated mass using LSR cranial equations. Combined-sex test sample (n=50), ordered by percentage of individuals estimated within 20% of known mass

	Directional Difference		Absolute Difference			
Cranial	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)	20% (%)	Wilcoxon T-value
BPOR	-0.5 (16.2)	-5.4 (21.2)	12.3 (10.4)	17.1 (13.5)	64.0	0.82
BIOR	-1.3 (17.1)	-6.8 (22.9)	12.8 (11.3)	18.1 (15.3)	64.0	0.6
FMA3	-1.2 (17.8)	-6.8 (22.9)	13.4 (11.6)	18.9 (14.3)	58.3	0.65
BORB	-1.1 (17.8)	-6.8 (23.6)	13.8 (11.2)	19.3 (14.9)	58.0	0.66
LFM	-0.2 (18.2)	-5.5 (22.8)	13.6 (11.8)	18.8 (13.7)	58.0	0.94
ORBA1	-0.8 (18.0)	-6.5 (23.6)	13.7 (11.6)	19.2 (15.0)	58.0	0.74
ORBA2	-0.9 (18.0)	-6.6 (23.6)	13.7 (11.6)	19.2 (15.0)	58.0	0.72
FMA1	-0.4 (17.8)	-5.8 (22.9)	13.6 (11.4)	18.9 (14.0)	58.0	0.87
FMA2	-0.5 (17.8)	-5.9 (22.9)	13.6 (11.4)	18.9 (14.0)	58.0	0.85
ORBA3	-2.4 (18.0)	-8.7 (23.8)	13.6 (11.8)	19.5 (15.9)	56.2	0.36
HORB	-1.1 (18.4)	-7.1 (24.2)	14.2 (11.7)	19.9 (15.3)	56.0	0.68
BFM	-0.9 (18.0)	-6.7 (23.7)	13.9 (11.2)	19.6 (14.7)	56.0	0.71

Table 4.16. Summary of difference between known and estimated mass using LSR postcranial equations. Combined-sex test sample (n=50), ordered by percentage of individuals estimated within 20% of known mass

	Directional Difference		Absolute Difference			
Postcranial	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)	20% (%)	Wilcoxon T-value
CA35	0.8 (11.3)	-1.6 (14.6)	8.9 (6.8)	11.9 (8.4)	82.0	0.62
CA65	-2.1 (12.5)	-5.9 (16.2)	10.2 (7.3)	14.1 (9.9)	80.0	0.24
MLSB50	1.0 (12.6)	-1.7 (16.0)	9.2 (8.6)	12.3 (10.2)	80.0	0.58
CA50	2.3 (11.7)	0.7 (14.4)	8.6 (8.2)	11.1 (8.9)	80.0	0.17
MLSB65	1.4 (13.4)	-1.4 (16.5)	9.9 (9.1)	13.0 (10.1)	78.0	0.46
CA80	1.1 (12.9)	-1.7 (16.5)	9.6 (8.5)	12.8 (10.3)	76.0	0.56
MLSB35	0.1 (13.9)	-3.5 (18.1)	10.9 (8.5)	14.7 (10.8)	74.0	0.95
FHB	1.3 (15.7)	-2.3 (19.4)	12.1 (9.8)	16.2 (10.7)	70.0	0.55
CA20	-2.8 (14.2)	-7.5(19.4)	11.7 (8.3)	16.5 (12.4)	70.0	0.17

	Directional Difference		Absolute Difference			
Postcranial	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)	20% (%)	Wilcoxon T-value
MLSB80	-1.2 (14.6)	-5.5 (18.8)	11.1 (9.4)	15.4 (12.0)	68.0	0.56
BIB	3.0 (16.8)	-0.5 (21.4)	12.7 (11.3)	16.8 (13.0)	64.0	0.21
MLSB20	-0.1 (16.1)	-4.6 (20.9)	12.6 (9.9)	17.2 (12.5)	62.0	0.97
FLM	-0.4 (17.0)	-5.2 (21.6)	13.2 (10.6)	18.1 (12.5)	60.0	0.88
MLCB65	-1.0 (18.4)	-7.0 (24.8)	14.4 (11.3)	20.4 (15.5)	58.0	0.69
MLCB50	-0.3 (18.8)	-6.0 (24.8)	14.5 (11.7)	20.2 (15.4)	58.0	0.92
FNB	-2.2 (16.1)	-7.2 (21.1)	13.2 (9.3)	18.4 (12.4)	56.0	0.35
MLCB80	-1.6 (18.5)	-7.9 (26.2)	14.6 (11.4)	20.6 (18.3)	54.0	0.54
MLCB35	-1.7 (18.5)	-8.1 (24.7)	14.6 (11.4)	20.6 (15.6)	54.0	0.51
MLCB20	-2.6 (18.4)	-9.3 (24.8)	14.7 (11.2)	21.0 (15.9)	54.0	0.31
I80	-101.1 (16.0)	-144.1 (42.3)	101.1 (16.0)	144.1 (42.3)	0.0	0.00
I65	-108.4 (16.3)	-153.6 (41.0)	108.4 (16.3)	153.6 (41.0)	0.0	0.00
I50	-150.7 (23.4)	-211.0 (48.4)	150.7 (23.4)	211.0 (48.4)	0.0	0.00
I35	-117.4 (17.4)	-166.1 (44.6)	117.4 (17.4)	166.1 (44.6)	0.0	0.00
I20	-109.7 (20.7)	-156.6 (49.4)	109.7 (20.7)	156.6 (49.4)	0.0	0.00

Female test sample: The best performing cranial variables in this sample were biorbital breadth (BIOR) and orbital area calculated as l x w (ORBA1) (Tables 4.17 and 4.18). These variables resulted in mean |PE|s of 17.8 and 19.2% respectively and estimated 56% of the sample within +/-20% of known mass. The least accurate cranial variable was orbital area, calculated as an ellipse (ORBA3). The equation for this variable returned a |PE| of 18.3%, and estimated 48% of the sample within +/-20% of known mass.

The best postcranial predictor of body mass in females was the index of cortical area at 80% of femur length (CA80), which returned a |PE| of 13.8% and estimated 80% of the sample within +/-20% of known mass. Mediolateral cortical breadth at 35% of femur length, as well as the five second moment of area indices, all returned |PE|s in excess of 100% and failed to estimate any individuals within +/-20% of known mass.

Table 4.17. Summary of difference between known and estimated mass using LSR cranial equations. Female test sample (n=25), ordered by percentage of individuals estimated within 20% of known mass

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

Table 4.18. Summary of difference between known and estimated mass using LSR postcranial equations. Female test sample (n=25), ordered by percentage of individuals estimated within 20% of known mass

	Directional Difference		Absolute Difference			
Postcranial	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)	20% (%)	Wilcoxon T-value
CA35	-3.6 (10.1)	-7.5 (16.4)	8.6 (6.2)	13.8 (11.3)	80.0	0.09
CA65	-3.6 (10.9	-7.3 (16.8)	10.0 (5.3)	15.4 (9.4)	76.0	0.11
MLSB50	-3.0 (11.0)	-6.3 (17.2)	8.9 (6.9)	13.8 (11.7)	76.0	0.18
CA50	-2.7 (11.5)	-7.0 (18.4)	9.6 (6.7)	15.4 (11.9)	72.0	0.26
MLSB65	-2.6 (10.3	-6.4 (16.4)	8.5 (6.1)	13.7 (10.8)	72.0	0.21
CA80	-4.8 (10.0)	-9.3 (16.0)	9.7 (5.3)	15.3 (10.0)	72.0	0.02
MLSB35	-2.3 (11.2)	-5.8 (17.8)	9.4 (6.3)	14.8 (11.2)	72.0	0.31
FHB	-3.2 (12.3)	-7.2 (19.7)	9.2 (8.6)	14.7 (14.8)	72.0	0.2
CA20	-3.0 (12.2)	-7.3(18.9)	10.1 (7.2)	15.9 (12.1)	68.0	0.24

	Directional Difference		Absolute Difference			
Postcranial	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)	20% (%)	Wilcoxon T-value
MLSB80	-5.4 (12.7)	-11.9 (21.0)	11.5 (7.5)	19.0 (14.6)	64.0	0.04
BIB	-3.4 (12.6)	-8.5 (19.7)	10.8 (7.0)	17.3 (12.3)	64.0	0.19
MLSB20	-4.5 (12.8)	-10.6 (21.2)	11.4 (7.0)	18.9 (13.9)	64.0	0.09
FLM	-4.4 (14.2)	-10.2 (23.7)	11.5 (9.2)	18.8 (17.3)	60.0	0.13
MLCB65	-3.3 (13.0)	-8.7 (20.8)	11.3 (6.9)	18.3 (12.6)	56.0	0.22
MLCB50	-3.1 (14.0)	-8.5 (22.1)	12.0 (7.5)	19.2 (13.4)	56.0	0.28
FNB	-2.4 (13.6)	-7.3 (21.5)	11.3 (7.6(	18.1 (13.2)	52.0	0.38
MLCB80	-3.8 (13.5)	-9.7 (21.2)	11.9 (7.0)	19.3 (12.6)	52.0	0.17
MLCB35	-7.0 (13.3)	-14.1 (22.5)	12.2 (8.5)	20.5 (16.6)	52.0	0.01
MLCB20	-141.6 (19.5)	-221.9 (54.5)	141.6 (19.5)	221.9 (54.5)	0.0	0
I80	-130.1 (19.2)	-203.0 (46.6)	130.1 (19.2)	203.0 (46.6)	0.0	0
I65	-194.0 (32.8)	-301.3 (67.3)	194.0 (32.8)	301.3 (67.3)	0.0	0
I50	-89.3 (14.4)	-143.3 (47.0)	89.3 (14.4)	143.3 (47.0)	0.0	0
I35	-140.4 (22.5)	-219.8 (55.6)	140.4 (22.5)	219.8 (55.6)	0.0	0
I20	-104. 1 (20.5)	-164.4 (50.1)	104.1 (20.5)	164.4 (50.1)	0.0	0

Male test sample: The best cranial predictor of body mass in males was biporionic breadth (BPOR) (Tables 4.19 and 4.20). This variable returned a |PE| of 15.9% and estimated 76% of the sample within +/-20% of known mass. The least accurate predictors were orbital height (HORB), foramen magnum breadth (BFM), orbital area as an ellipse (ORBA2) and orbital area as b x h (ORBA1). These variables resulted in |PE|s between 17.6-19.2% and estimated 60% of the sample within +/-20% of known mass.

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

Table 4.19. Summary of difference between known and estimated mass using LSR cranial equations. Male test sample (n=25), ordered by percentage of individuals estimated within 20% of known mass

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

Table 4.20. Summary of difference between known and estimated mass using LSR postcranial equations. Male test sample (n=25), ordered by percentage of individuals estimated within 20% of known mass

	Directional Difference		Absolute Difference			
Postcranial	**Raw diff (kg) Mean (SD)**	**PE Mean (SD)**	**Raw diff (kg) Mean (SD)**	**|PE| Mean (SD)**	**20% (%)**	**Wilcoxon T-value**
CA35	9.0 (19.6)	6.6 (20.3)	15.8 (14.4)	17.6 (11.7)	65	0.0
CA65	1.7 (20.5)	-3.2 (24.5)	15.7 (12.9)	19.4 (14.9)	64	0.7
MLSB50	7.8 (21.1)	4.4 (23.7)	16.9 (14.6)	19.6 (13.5)	64	0.1
CA50	9.2 (19.9)	6.6 (21.2)	16.4 (14.4)	18.4 (11.9)	64	0.0
MLSB65	0.2 (19.6)	-4.9 (23.5)	15.3 (11.9)	19.1 (14.1)	60	1.0
CA80	5.2 (19.9)	1.5 (23.0)	15.5 (13.1)	18.4 (13.3)	60	0.2
MLSB35	7.9 (20.2)	4.8 (22.2)	16.2 (14.2)	18.6 (12.5)	60	0.1
FHB	1.9 (19.7)	-2.7 (23.0)	14.9 (12.6)	18.3 (13.7)	56	0.6
CA20	2.4 (19.9)	-2.2 (23.3)	15.1 (12.8)	18.4 (13.9)	56	0.6

	Directional Difference		Absolute Difference			
Postcranial	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)	20% (%)	Wilcoxon T-value
MLSB80	6.0 (21.1)	2.0 (24.3)	16.7 (13.8)	19.7 (13.7)	56	0.2
BIB	10.0 (21.2)	7.3 (23.0)	17.5 (15.4)	19.7 (13.3)	56	0.0
MLSB20	10.2 (19.6)	8.1 (20.1)	16.1 (14.9)	17.7 (12.1)	56	0.0
FLM	11.3 (19.9)	9.4 (20.5)	16.6 (15.6)	18.2 (12.9)	56	0.0
MLCB65	12.4 (19.9)	11.0 (20.2)	17.2 (15.8)	18.8 (12.9)	56	0.0
MLCB50	0.4 (19.2)	-4.4 (22.8)	15.0 (11.7)	18.6 (13.4)	52	0.9
FNB	7.8 (20.3)	4.70 (22.3)	16.4 (14.0)	18.8 (12.3)	52	0.1
MLCB80	9.8(22.0)	6.8 (23.2)	18.0 (15.8)	20.2 (12.7)	52	0.0
MLCB35	16.3 (21.6)	15.1 (21.5)	20.8 (17.1)	22.8 (12.6)	48	0.0
MLCB20	12.5 (22.3)	10.1 (23.3)	19.2 (16.6)	21.5 (13.1)	40	0.0
I80	-33.9 (22.3)	-48.7 (37.1)	36.3 (17.9)	50.3 (34.7)	16	0.0
I65	-47.9 (21.0)	-66.0 (38.6)	49.3 (17.6)	66.9 (36.9)	4	0.0
I50	-59.8 (21.0)	-80.3 (39.3)	60.4 (19.4)	80.7 (38.5)	4	0.0
I35	-80.4 (23.0)	-106.6 (46.9)	80.4 (23.0)	106.6 (46.9)	4	0.0
I20	-52.0 (23.2)	-71.7 (42.8)	53.2 (20.1)	72.5 (41.3)	4	0.0

4.4. Discussion

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

The results show that six of the 12 cranial equations can be considered valid in the combined-sex sample. Fourteen of the 24 postcranial equations met the criteria for acceptance in this group. In the female-only sample, all but two cranial equation

(ORBA3) and 14 of 24 postcranial equations were valid. In the male-only sample, all but one cranial equation met the criteria, while 11 of the postcranial equations were acceptable predictors. Thus, the majority of the equations can be considered reliable estimators of mass, according to the assessment criteria.

Table 4.21 summarizes a comparison between the results of our equations and those of previous studies as tested against our combined-sex test sample (n=50). The test employed the three "best" cranial predictors from Aiello and Wood (1994) (orbital area, orbital height and biporionic breadth) and the femoral head breadth equations provided by Ruff et al. (1991), McHenry (1992), Grine et al. (1995), and Ruff et al. (2012). Aiello and Wood's (1994) equations were used rather than those of Kappelman (1996) and Spocter and Manger (2007) because they produced lower levels of error in a previous assessment (Elliott et al., 2014). The four combined-sex FHB equations were included because they are all regularly applied (e.g. Trinkaus et al., 2014). We used only the combined-sex sample to ensure a large sample size and because sex is often difficult to attribute in fragmentary skeletal remains.

Table 4.21. Comparison of the present results with those of previously published regression equations (combined-sex test sample, n=50)

	estimate (kg) mean (SD)	raw diff (kg) mean (SD)	|PE| (SD)	20%
Known Mass	74.6		--	--
ORBA1 (Aiello and Wood 1994)	53.5 (10.6)	21.1 (19.2)	28.7 (15.0)	28
ORBA1 (present study)	75.5 (3.2)	0.9 (18.0)	19.2 (15.0)	58
HORB (Aiello and Wood 1994)	47.1 (11.6)	27.5 (20.9)	35.7 (17.7)	26
HORB present study	75.7 (1.1)	-1.1 (18.4)	19.9 (15.3)	56
BPOR (Aiello and Wood 1994)	69.7 (13.0)	4.9 (15.7)	15.6 (12.4)	70
BPOR present study	75.1 (5.6)	-0.5 (16.2)	17.1 (13.5)	64
FHB-1 (Ruff et al. 1991)	78.1 (8.3)	-3.5 (15.6)	18.5 (12.7)	62
FHB-2 (McHenry, 1992)	66.7 (8.7)	7.9 (15.6)	14.9 (11.4)	68
FHB-3 (Grine et al. 1995)	71.5 (8.8)	3.1 (15.6)	15.4 (10.5)	68
FHB-4 (Ruff et al. 2012)	67.8 (8.9)	6.8 (15.6)	14.8 (11.2)	68
FHB (Present study)	73.3 (7.9)	1.3 (15.7)	16.2 (10.7)	70

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

Given that the majority of the equations we generated met the criteria for accuracy and generally outperformed the best equations in the literature, the study achieved its goal of improving the estimation of body mass from skeletal remains. However, there are reasons to temper this conclusion.

One important issue is that the equations were generated and tested under ideal conditions: both the calibration and test samples comprised individuals of known body mass and sex, and the test sample was drawn from the same population as the calibration sample. Thus, the results represent a "best-case-scenario" and must be considered to be the upper limit of accuracy for estimating body mass from skeletal remains. Despite this, failure rates were surprisingly high and the body masses of many specimens were not estimated within 20% of their actual body masses. To reiterate, in the combined-sex sample, six of the 12 cranial variables failed to meet the |PE| criterion and none estimated more than 64% of the sample within 20% of their known mass. Ten of the 24 postcranial variables failed to meet the |PE| criterion and only four variables (CA35, CA65, MLSB50 and CA50) estimated 80% or more individuals within 20% of their known mass. This suggests that the utility of the standard, regression-based approach to estimating body mass from skeletal remains is much more limited than the field has appreciated. Even under ideal conditions, body mass estimates are not likely to be very accurate, and any deviation in terms of incomplete or distorted elements, sex uncertainty or proportional differences between the reference and target groups, will almost certainly result in greater error. As a result, estimating body mass using any regression equation must be undertaken very cautiously and the resulting masses considered ball-park figures at best.

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

Several expectations regarding postcranial variables were also not met. For example, femoral head breadth was not among the top five postcranial predictors for any of the samples and none of the FHB equations estimated more than 70% of any of the three test samples within +/-20% of known mass. This was surprising since femoral head breadth is the most widely-used skeletal variable for estimating body mass (Ruff et al., 1991; McHenry, 1992; Grine et al., 1996; Ruff, 2010; Ruff et al., 2012), in part because proximal femora are more common in the fossil record and the relevant measurements are easy to take (Ruff et al., 1997). Femoral head breadth has also been argued to be more appropriate for body mass estimation than other areas of the femur because it is less responsive to external influences like environmental stresses and activity (Ruff et al., 1997). Habitual activity in particular, is thought to influence femoral cross-sectional dimensions strongly (Ruff and Hayes, 1983; Ruff et al., 1984; Trinkaus et al., 1991; Trinkaus and Ruff, 1999). Despite this, several shaft dimensions produced lower rates of error than femoral head breadth in the present study. Consequently, these results suggest that FHB may not be as reliable an estimator of mass as it is usually assumed to be.

Although the extent to which diaphyseal dimensions are influenced by activity, environment, body mass or some combination of these factors continues to be poorly understood (Pearson and Lieberman, 2004; Pearson et al., 2008), our results suggest that shaft measurements should be investigated more thoroughly for their ability to

estimate individual mass reliably. As noted, mid-shaft dimensions, particularly cortical area indices and medio-lateral shaft breadths, consistently performed better than FHB. While obtaining cortical dimensions was a difficult task in the past as it required the use of two-dimensional radiography or physical sectioning of the bone (e.g. Ruff and Hayes, 1983), technologies like CT and MRI are becoming more accessible for anthropological research and offer the potential for accurate and non-destructive ways of accessing these data (Thali et al., 2003). Consequently, further research in this area should be pursued.

More surprising still, the results did not consistently support expectations regarding the relative reliability of cranial versus postcranial variables. Although functional relationships are not a prerequisite for good predictability (Smith, 2002), most researchers argue that postcranial features will estimate mass better than cranial features because they bear the body's weight (e.g. Jungers, 1990; Ruff et al., 1991). Our results suggest that this assumption needs to be examined more closely. Overall, the equations derived from the postcranial variables estimated mass more accurately than the cranial variables in the female and combined-sex samples (Table 4.22). In males, however, the postcranial variables were generally less accurate than the cranial variables. This suggests that the relationship between lower limb morphology and mass may be different in males than in females. While this may be due to variations in activity or muscle mass, as noted earlier, the relationship between skeletal morphology, activity and muscling continues to be a complex problem (Stirland, 1998; Weiss et al 2010; Takigawa, 2014). Although accurate data regarding activity patterns and muscle mass are difficult to obtain outside specialized research settings (Kim et al., 2002), the results obtained here argue strongly in favour of further research in this area.

Table 4.22. Comparison of the 20% criterion results for cranial and postcranial variables

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

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

By the same token, lower correlations did not necessarily indicate poorer predictability either. For example, in the combined-sex sample, the correlation coefficient between femoral neck breadth and mass was 0.43 ($p<0.01$) and 56% of the sample was estimated within +/-20% of known mass. In contrast, femoral head breadth (FHB) had a lower correlation ($r=0.40$, $p<0.01$) but estimated more of the sample (70%) within +/-20% of known mass. Interestingly, these correlations are similar to those in Ruff et al. (1991), where femoral neck breadth and femoral head breadth correlated with current weight in the combined-sex sample at $r=0.53$ and $r=0.49$ respectively. Although interspecific correlation coefficients are always expected to be higher than intraspecific ones (Smith, 2002), the lower correlation coefficient would argue against femoral head breadth as the more appropriate estimator of mass. Regardless, it is clear that correlation coefficients may not be a reliable means of assessing the predictive competence of an equation. While this point has been made before (Smith, 1984; 1985) it appears to have had little impact on practice (e.g. de Groot and Humphrey, 2011; Eller et al., 2014; Jung et al. 2014).

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

To explore these issues further, we examined the relationship between SEEs, correlation coefficients, absolute prediction errors and the percentage of individuals who are estimated within +/-20% of their known mass. Table 4.23 shows that although there is a reasonably good (inverse) relationship between SEEs and the correlation coefficients for a given variable, neither statistic is a good indicator of estimation competence as evaluated by the percentage difference from known mass and the "+/-20% criterion". Thus, in the absence of a known mass sample, neither the correlation coefficient of the variable, nor the SEE of the regression equation are adequate to determine the accuracy of the resulting estimate. Consequently, in order to ensure a regression equation is reliable, one must start with accurately measured variables combined with associated individual body masses. The resulting equation should be tested on a known, independent sample, ideally drawn from the same population as the reference group (Giancristofaro and Salmaso, 2007). Subsequent validation on known-mass samples from other populations would then provide additional confidence and broader applicability. But without these steps, equations cannot be adequately assessed for their ability to estimate mass in unknown specimens. Even under these conditions, only broad estimate ranges can be expected and any inferences drawn from them must be made with caution.

Table 4.23. Comparison of correlation with SEE, combined-sex calibration sample (n=203)

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

4.5. Conclusion

The results of this study support Smith's (2002:271) contention that body mass estimation is not the "simple matter" some researchers have perceived to be. The majority of the cranial and postcranial variables tested met the criteria for acceptance as estimators of mass. In addition, most of the equations returned lower rates of error than previously published equations for the same variables. However, the acceptance criteria used in the present study were lenient and the improvements over earlier studies were modest. Given the vagaries of taphonomy, uncertainties of sex attribution in fragmentary skeletal remains and the difficulty (or worse, impossibility) of ensuring a reference sample is appropriate for the target specimen, these results suggests that body mass estimation is fraught with more uncertainty than most applications acknowledge. In addition, the attempt to derive more accurate regression equations for estimating body mass revealed other problems. Specifically, the variables currently favoured for body mass estimation may not be the most reliable. In addition, variables with a functional association to mass (e.g. femoral head breadth) were not consistently better predictors than those without (e.g. orbital height) and the criteria currently employed to evaluate predictive competence did not assure accuracy. While some of the variables tested here show promise as predictors of mass, further research using large documented samples needs to be undertaken to address the issues that have been identified.

Chapter 5.

General Discussion

The results of these three studies are not encouraging for the estimation of body mass from skeletal remains as it is currently undertaken in biological anthropology. However, several limitations of the present research must be considered before the described results are accepted. The first relates to the nature of the reference sample used. During the peer-review process for the first paper, our use of individual modern humans to test regression equations derived from inter-specific mean data was criticized (Ruff, pers. comm.). Because the relationship between body mass and a skeletal variable inevitably differs between-species as compared to within-species, inter-specific regressions are assumed to be applicable only to estimating the average mass of a fossil species (Gingerich, 1977). However, this is not how the equations are used in practice and our decision was justified for all three studies for several reasons.

First, an exclusively modern human sample was selected because it was large, fully documented with associated skeletal measurements and body masses, and independent of all the samples previously used to generate the regression equations. As a result, it was an excellent choice for testing the postcranial equations, which were also based on human samples, as well as for the third study, which derived new regressions using a more robust sample than had been employed before. It was also appropriate for the first study because the goal was to assess published cranial equations using a known-mass sample, something that had not been done before. Since modern humans comprised two out of the 12 hominoid groups in Aiello and Wood (1994), two of 18 groups in Kappelman (1996) and one fifth of the groups in Spocter and Manger (2007), the analyses did not extrapolate beyond the reference range (Auerbach and Ruff, 2004) or violate the statistical requirement for the test specimens to be members of one of the same populations used to derive the regressions (Smith, 2002; Wood, 2011). While it would have been informative to include non-human primates in the analyses, as noted

previously, such samples are rare, and those that do exist are likely to have been included in the material used to generate the regression equations in the first place. Also, there are currently no equivalent CT datasets for non-human primates that include individual associated body masses.

Importantly, the decision to use individual data rather than sample means provided more fine-grained analyses than would have been achieved otherwise. For the second and third studies, this was critical because the equations have been argued to be appropriate for estimating individual mass (Ruff et al., 1991) and are routinely used to do so (Ruff, 2010; Trinkaus et al., 2014). For the first study, individual data were used in order to explicitly test current anthropological practice. As described earlier, the three sets of cranial equations employed species or sex specific mean values for a variety of extant primates. As such, the assumption is that the resulting body mass represents an average species mass only, and is not used to evaluate differences between individual fossil specimens. However, this is not the case. Specifically, all the original cranial studies (Aiello and Wood, 1994; Kappelman, 1996; Spocter and Manger, 2007) as well as others (e.g. Kordos and Begun, 2001; Rightmire, 2004), have estimated masses for multiple individuals within a species from the mean interspecific regressions. Furthermore, the resulting masses were not necessarily combined into a single average for the taxon or interpreted that way. Consequently, the present research tested the equations as they would be applied to single fossil specimens, thus replicating the way in which cranial measurements are actually used to estimate body mass in biological anthropology. The decision was also supported by supplementary analyses carried out on mean data for multiple sub-sets of the sample, as described in the first paper.

Another potentially confounding factor in the present research relates to the body condition of the sample individuals. On the grounds that extreme outliers may negatively bias the predictions, many studies that employ modern human samples exclude emaciated and obese individuals from their analyses (Holloway, 1980; Sciulli and Blatt, 2008; Daneshvari, 2011). Obese individuals, in particular, are removed on the grounds that obesity is a relatively modern phenomenon and past groups are unlikely to have carried excess weight. In fact, this was the logic behind the recommendation to use a downward correction factor of approximately 10% when estimating the mass of earlier human samples and fossil hominins with regressions derived from modern groups (Ruff

et al., 1991; Ruff, 2010). Despite this, I chose not to exclude individuals whose body masses fell above or below the World Health Organization definition (WHO, 2000) of "normal" (BMI 18.5-25) for several reasons. First, including the full range of BMIs provided the opportunity to explicitly test the effect of body condition on estimate accuracy. To do this, analyses were carried out on BMI restricted samples for both the cranial and postcranial equations. The results for the cranial equations are provided in Appendix Tables 6-9. Appendix table A16 provides the summary data for the postcranial variables in the BMI-restricted sample, while table A17 summarizes the postcranial results. In both cases, estimation accuracy improved for some equations, but not for all and not consistently. Since there is still some question as to whether or not past groups were more heavily muscled than current populations (Pearson et al., 2008) and muscle weighs approximately 13% heavier than fat (Snyder et al., 1975; Janssen et al., 2000; Kuczmarski and Flegal, 2000), the exclusion of obese individuals in reference samples (and the 10% downward correction) may be unwarranted. Lastly, if one of the potential uses of body mass estimation is to discriminate one individual from a list of candidates, as in a forensic context (Stubblefield, 2003; Moore and Schaefer, 2011; Lorkiewicz-Muszyńska et al., 2013), then the ability to identify extremes of body mass may actually be more informative. This issue requires further investigation.

The broad age range of the sample (18-90 yrs) was also a potentially confounding factor in the current research. Body mass has been argued to increase with age, particularly in females and after the fifth decade, as a result of increased fat accumulation (Holloway, 1980; Ruff et al., 1991, 2005). In addition, past populations are less likely to have lived into very old age (Robson and Wood, 2008), making the inclusion of older individuals in reference samples potentially unnecessary. However, body mass may also decrease with advanced aged (>60) due to inactivity, cachexia and sarcopenia (Seidell et al., 2000; Perissinotto et al., 2002). Age is also extremely difficult to assess once adulthood is reached, particularly in populations whose growth and senescent trajectories may not be the same as modern humans (Dean et al., 2014). Stature also decreases with age, as a result of disc compression, fractures, and postural changes (Cline et al., 1989). This factor may be particularly relevant for the morphometric postcranial equations and for females, as the effect may be exaggerated due to their higher susceptibility to osteoporosis (Pothiwala et al., 2006). Accordingly, the extent to which body mass predictions could vary with age is not clear.

To explore the effect of age on estimation accuracy in the present research, I conducted two additional sets of analyses. First, correlation coefficients for the absolute PPEs for the published regression equations were plotted against age (Appendix Tables 18 and 19). Aiello and Wood's (1994) LSR equations were used for the cranial analyses, while both the mechanical and morphometric equations were used for the postcranial analyses. Absolute differences were used in order to assess the magnitude of the relationship, rather than the direction. For most of the cranial equations, correlations are slightly positive, indicating that prediction errors get somewhat larger with age. For the postcranial equations, a similar pattern exists for females and the combined-sex samples, but not consistently for males. These results suggest that age-related body mass differences between females and males may be an important consideration. However, the effect is small and further research is needed before age adjustments could be recommended. A similar test with stature found an increase in prediction error (overestimation) with age (Ruff et al., 2012) and a more exaggerated response in females. However, as in the current study, the effect was small in both sexes and Ruff et al. (2012) concluded that an age adjustment was unnecessary.

A second set of analyses used the new regression equations and the same test sample as in paper 3, but restricted the test sample to individuals between 18 and 60 years (n=46). For most of the cranial and postcranial variables, absolute percent prediction errors did not change significantly when 'elderly' (>60 yrs) individuals were removed (Appendix Tables 20 and 21). Several variables (femoral head breadth, I80, I65, I50) showed a small drop in |PPE|, but others (FNB, CA35), showed a slight increase when age was restricted. The reasons for these results are unclear, but may be worth investigating.

In sum, although there are several factors that may lead to greater degrees of inaccuracy when regression equations for estimating body mass are applied to fossil, archaeological or modern groups, the present research attempted to mitigate these issues as much as possible. In addition, the methods and evaluation criteria were deliberately lenient, allowing for the greatest opportunity for success given the application of the various equations. This is a critical point: since this research represents the best case scenario for estimating body mass in biological anthropology,

any deviation from the conditions adhered to here will almost certainly result in greater error.

The potential impact of the current research on biological anthropology has been outlined in the individual studies. However, there are two broader implications that should be stressed. First, it is clear that existing equations for estimating mass must be treated more cautiously than they have been to date. Despite using test individuals from the same population as the reference sample, error rates in the three studies reported here were surprisingly high. As a result, previously published equations derived from less robust samples are likely to produce even higher rates of error because they were derived from smaller samples, used unassociated material and did not necessarily employ known body masses. Consequently, these equations should be used carefully, regardless of the target specimens.

It is also clear that the body masses of fossil hominins and other specimens that have been estimated using these equations need to be reviewed. For example, the Tyrolean "ice-man" whose body was discovered in the Swiss Alps in 1991 (Seidler et al., 1992) has a femoral head diameter of 44.3 mm (Ruff et al., 2006). Using an average of the three published equations (Ruff et al., 1991; McHenry, 1992; Grine et al., 1995), as recommended by Auerbach and Ruff (2004), this individual was estimated at 61 kg (Ruff et al., 2006). However, using the regression equation for FHB derived from our modern Swiss sample, "Oetzi's" body mass would be 65.1 kg. While this is still well within the normal human range, a difference of almost 7% could have a significant impact on interpretations of this individual's robusticity and activity patterns (Ruff et al., 2006). A more extreme example is evident when considering OH5, a large male attributed to *Paranthropus boisei* (Leakey, 1959). Using Aiello and Wood's (1994) equations for orbital height and orbital area, this individual would be estimated at 40.0 or 67.0 kg respectively. However, on the basis of the only equation that satisfied both of the criteria for reliability in our first study (biporionic breadth), this same specimen would be estimated at 122.2 kg. Clearly, if the latter is more accurate, the current interpretations of this species biology and behaviour require revision.

The results obtained in the three studies reported here raise the question of how best to proceed in future. As a preliminary step, it is imperative that the most robust and

comprehensive samples be used when testing existing methods or attempting to develop new ones. Without employing reference material comprised of known values, it is impossible to define the true nature of the relationship between a skeletal feature and body mass. Consequently, any subsequent prediction made on the basis of that relationship will necessarily be even more imprecise. In addition, if the relationship is not verified in an extant group, it is not possible to determine if, or to what extent, the relationship holds in past species (Smith, 1996).

To this end, biological anthropologists with appropriately robust samples should be encouraged to share their datasets, and be as detailed and explicit about their collection methods and procedures as possible. There are a number of promising examples in this regard, some of which encourage contributions on an ongoing basis (e.g. Jantz and Ousley, 2005). To this end, the data for the present study will be made available upon request. However, a better approach would be to develop a single large, repository of material that could be available for diverse avenues of research. While there are significant practical, financial and legal challenges to such a task, this should not stop researchers from pursuing this goal.

The acquisition of similar datasets for non-human primates must also be pursued. Well-documented non-human primate collections have always been rare (Kappelman, 1996), but ethical and conservation issues have exacerbated the problem by preventing the replacement of existing material or the development of new collections (Holloway, 1980). However, zoo, research and wildlife facilities continue to house numerous primate species and many maintain detailed records of the individuals under their care (AZA, 2014). While there are national and international laws regulating the distribution of animals or their parts after death for ethical and public health reasons (IUCN, 2014), CT technology offers the ability to obtain the necessary skeletal and biological information without compromising the integrity of the remains. Since the resulting data would then be permanently accessible to researchers in multiple disciplines, even remotely, zoo and wildlife facilities might be encouraged to enter into such collaborations in order to maximize resources and raise their own profiles. These relationships should be sought out and supported.

The incorporation of new technology should also be strongly encouraged. The use of CT has a long history in anthropology (Jungers and Minns, 1979), but the possibilities for research truly exploded with the development of three dimensional visualization tools and reconstruction software (e.g. De Greef and Willems, 2005; Balzeau et al., 2010). As powerful as CT is, it is not the only technology with potential benefits for those interested in estimating body mass from skeletal remains. Magnetic Resonance Imaging (MRI) and Dual-energy X-ray Absorptiometry (DXA) also hold significant potential as research tools. While originally designed for soft-tissue analyses (Lauterbur, 1973), MRI is now capable of detailed and accurate skeletal assessments (Woodhead et al., 2001; Anastasi et al., 2009; Spahic et al., 2009). In addition, MRI is does not require the same level of radiation exposure that CT does. Consequently, large-scale in-vivo studies of non-pathological individuals are possible, even in juveniles (Vannucci et al., 2011). In addition, whole body MRI scans are now being offered to the public for prophylactic purposes (CMI, 2014). While there may be selection biases in terms of who has the interest and financial means to undertake such scanning, these data offer the unique ability to review anatomical details on large numbers of non-pathological individuals in more detail than traditional radiographs and with lower risk. DXA is also of interest, because it allows for the analysis and quantification of bone mineral density, an important measure of biomechanical load in skeletal studies (Lam and Pearson, 2005). DXA also has the ability to measure total body composition, including lean muscle and fat and is being increasingly used in anthropological applications (Whitmarsh et al., 2010; Castillo et al., 2011). Both MRI and DXA hold considerable potential for research on body mass estimation from skeletal remains and should be taken advantage of whenever possible.

Future research may also benefit from taking a different approach to the problem. In biological anthropology, body mass is rarely the feature of interest per se. Instead, the goal is usually to use an estimate to establish some other aspect of an individual or species' biology or behaviour that corresponds to size. As Smith (1996) points out, a host of ecological, physiological and behavioural traits have been "determined" this way. However, as he also suggests, for biological anthropologists seeking to predict such life history characteristics, it may be more fruitful to eliminate the intermediate variable (body mass) and go straight to an analysis of the relationship between the skeletal feature and the life history trait. Although there is no functional reason for the two to correspond, this

is not necessary for good predictability (Smith, 2002) and this approach would eliminate the error inherent in making a two-step prediction. It would also encourage more circumspection when interpreting the results. While this is certainly an unconventional proposal and one that has not been taken up enthusiastically, it may be worth considering.

Different statistical approaches may also be constructive. Alternatives to the standard single-variable technique of regression of mass on a skeletal measure have been proposed several times (Konigsberg et al., 1998; Hens et al., 2000). Multivariate regression analyses have also been employed in the past (e.g. Jungers, 1990) and have certainly been encouraged (Smith, 1996, 2002). Despite this, univariate "inverse calibration" methods still dominate body mass analyses (Pomeroy and Stock, 2012), presumably because most skeletal material still consists of isolated elements. This should not however, prevent the exploration of novel multivariate methods as a way to ensure more options are available. A number of studies have started pursuing other options (Daneshvari, 2011; Moore and Schaefer, 2011; Konigsberg and Frankenberg, 2013; Uhl et al 2013) with promising results. Ideally, whatever method is employed, it should use a robust, documented sample and a combination of cranial and postcranial material. Such methods would then be available when associated skeletal material is found.

Chapter 6.

Conclusions

This research had two primary goals. The first was to test existing methods for estimating body mass from skeletal material. The second was to derive new regression equations for improving this task. In both cases, the approach was to use a large sample, directly measured variables and individual, associated body masses. As a result, I expected the sample to provide the most robust test of existing methods and be the best way to capture the true nature of the relationship between skeletal morphology and mass accurately. The results were mixed.

Previously published equations for estimating body mass from cranial variables produced high rates of error in a sample of modern humans. Body mass estimates varied between studies, as did the predictive competence of the different equations for the same variables. In particular, two variables (orbital height and orbital area) that have been argued to be good predictors in the past, did not estimate mass reliably in the test sample. Thus, the continued use of these cranial equations cannot be recommended.

Published postcranial equations estimated more of the test sample accurately. However, predictive competence was not consistent and variables that were expected to estimate mass reliably did not always do so. In addition, several assumptions regarding the way the equations are applied in biological anthropology were not supported. Specifically, morphometric equations are not more reliable than mechanical equations, even though they are based on a larger, more diverse sample of modern humans. The more recently published morphometric equations were not more appropriate for high latitude groups than earlier equations. However, averaging sex-specific morphometric equations when the sex of the specimen is unknown is acceptable. With respect to the mechanical method, there is very little advantage to averaging the results of multiple equations and sex-specific equations are not consistently more accurate that those derived for combined-sexes. Furthermore, equations designed for particular body types

did not necessarily estimate mass in accordance with expectations for their target groups.

The predictive ability of previously published regression equations may have been compromised by small reference samples, the use of indirect measures for key variables and the inability to use associated skeletal material and body mass data. Consequently, I assumed that employing a large calibration sample of directly measured variables and associated, individual data, would significantly improve the accuracy of body mass estimates. This was only partially the case. The present regression equations estimated mass more accurately on a well-matched test sample than most previously published equations. However, for both cranial and postcranial variables, estimation accuracy was lower than expected and not always consistent. Although it may be advantageous to employ these equations rather than existing ones when estimating mass in fossil hominins and archaeological or modern human groups, the results suggest that all body mass estimates are subject to significant error.

Further analyses involving large samples of known individuals are needed to resolve the issues identified in these studies. Incorporating comparable data for non-human primates into these analyses would be ideal. As appropriate to the research question and resources, the use of medical technologies like CT, MRI and DXA should be encouraged. Finally, new statistical approaches should be investigated to help improve predictive ability. With refinements such as these, it may be possible to increase the accuracy of body mass estimates from skeletal material. Until then, existing mass estimates must be viewed critically and all equations used judiciously.

References

Adams GL, Gansky SA, Miller AJ, Harrell Jr WE, Hatcher DC. 2004. Comparison between traditional 2-dimensional cephalometry and a 3-dimensional approach on human dry skulls. American Journal of Orthodontics and Dentofacial Orthopedics. 126(4):397-409.

Aghayev E, Staub L, Dirnhofer R, Ambrose T, Jackowski C, Yen K, Bolliger S, Christe A, Roeder C, Aebi M, et al. 2008. Virtopsy - the concept of a centralized database in forensic medicine for analysis and comparison of radiological and autopsy data. Journal of Forensic Legal Medicine. 15(3):135-140.

Agostini GM, Ross AH. 2011. The Effect of Weight on the Femur: A Cross-Sectional Analysis. Journal of Forensic Sciences. 56(2):339-343.

Aiello LC. 1992. Allometry and the analysis of size and shape in human evolution. Journal of Human Evolution. 22(2):127-147.

Aiello LC, Key C. 2002. Energetic consequences of being a Homo erectus female. American Journal of Human Biology. 14:551-565.

Aiello LC, Wood B. 1994. Cranial variables as predictors of hominine body mass. American Journal of Physical Anthropology. 95(4):409-426.

American Association of Zoos and Aquariums. 2014. Animal Care and Management. https://www.aza.org/animal-programs.

Anastasi G, Cutroneo G, Bruschetta D, Trimarchi F, Ielitro G, Cammaroto S, Duca A, Bramanti P, Favaloro A, Vaccarino G et al. 2009. Three-dimensional volume rendering of the ankle based on magnetic resonance images enables the generation of images comparable to real anatomy. Journal of Anatomy. 215(5):592-599.

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

Auerbach BM, Ruff CB. 2004. Human body mass estimation: A comparison of “morphometric” and “mechanical” methods. American Journal of Physical Anthropology. 125(4):331-342.

Balzeau A, Crevecoeur I, Rougier H, Froment A, Gilissen E, Grimaud-Hervè D, Mennecier P, Semal P. 2010. Applications of imaging methodologies to paleoanthropology: Beneficial results relating to the preservation, management and development of collections. Comptes rendus de l'Académie des Sciences. 9(6-7):265-275.

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

Bassed R, Drummer O, Briggs C, Valenzuela A. 2011. Age estimation and the medial clavicular epiphysis: analysis of the age of majority in an Australian population using computed tomography. Forensic Science, Medicine, and Pathology. 7(2):148-154.

Bayomi DJ, Tate RB. 2008. Ability and accuracy of long-term weight recall by elderly males: the Manitoba follow-up study. Annals of Epidemiology. 18:36-42.

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

Byard RW. 2012. The complex spectrum of forensic issues arising from obesity. Forensic Science, Medicine and Pathology. 8:402-413.

Beckmann EC. 2006. CT scanning the early days. British Journal of Radiology. 79(937):5-8.

Blumenberg B. 1984. Allometry and evolution of tertiary hominoids. Journal of Human Evolution. 13(8):613-676.

Byard RW. 2012. The complex spectrum of forensic issues arising from obesity. Forensic Science, Medicine, and Pathology. 8(4):402-413.

Calder WA. 1984. Size, function, and life history. Cambridge, Massachusetts and London, England: Harvard University Press. 431 p.

Caspari R. 2003. From Types to Populations: A century of race, physical anthropology and the American Anthropological Association. American Anthropologist. 105(1):65-76.

Castillo RF, Lopez RM. 2011. Assessment of age and sex by means of DXA bone densitometry: Application in forensic anthropology. Forensic Science International. 209(1), 53-58.

Cavalcanti M, Rocha S, Vannier M. 2004. Craniofacial measurements based on 3D-CT volume rendering: implications for clinical applications. Dentomaxillofacial Radiology. 33(3):170-176.

Center for Medical Imaging. 2014. Patient information on MRI. http://www.cmiradiology.com/mri.php

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. p 319-346.

Cline MG, Meredith KE, Boyer JT, Burrows B. 1989. Decline of height with age in adults in a general population sample: estimating maximum height and distinguishing birth cohort effects from actual loss of stature with aging. Human Biology, 415-425.

Cohen MN, Crane-Kramer GMM. 2007. Ancient health: Skeletal indicators of agricultural and economic intensification. Gainesville: University Press of Florida. 464 p.

Collard M, Cross, AG. (in press). Thermoregulation in Homo erectus and the Neanderthals: a reassessment using a three-dimensional, segmented model. In: Papers in Honour of Yoel Rak. E. Hovers and A. Marom (eds). Berlin: Springer-Verglag.

Conroy G. 1987. Problems of body-weight estimation in fossil primates. International Journal of Primatology. 8(2):115-137.

Cross AG, Collard M. 2011. Estimating Surface Area in Early Hominins. PloS one. 6(1).

Dagosto M, Terranova C. 1992. Estimating the body size of eocene primates: A comparison of results from dental and postcranial variables. International Journal of Primatology. 13(3):307-344.

Damuth J, MacFadden B. 1990. Body size in mammalian paleobiology. Cambridge, UK: Cambridge Univ. Press. 412 p.

Daneshvari S. 2011. Predicting Body Mass from the Skeleton with an Application to the Georgia Coast. Unpublished PhD thesis. Albuquerque: University of New Mexico.

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. Journal of Forensic Sciences. 5:1107-1114.

De Greef S, Willems G. 2005. Three-dimensional Cranio-Facial Reconstruction in Forensic Identification: Latest Progress and New Tendencies in the 21st Century. Journal of Forensic Sciences. 50(1):1-6.

De Groote I, Humphrey LT. 2011. Body mass and stature estimation based on the first metatarsal in humans. American Journal of Physical Anthropology, 144(4), 625-632.

Dean MC, Liversidge HM, Elamin F. 2014. Combining radiographic and histological data for dental development to compare growth in the past and the present. Annals of Human Biology. 41(4):336-347.

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. New York: American Museum of Natural History. 159 p.

DeSilva JM. 2011. A shift toward birthing relatively large infants early in human evolution. Proceedings of the National Academy of Sciences. 108(3):1022-1027.

DeSilva JM, Lesnik JJ. 2008. Brain size at birth throughout human evolution: a new method for estimating neonatal brain size in hominins. Journal of Human Evolution. 55:1064-1074.

Duncan WS. 1883. The Fossil Antecedents of Man, and where to Discover Them: London: Harrison and Sons. 54 p.

Dupont WD, Plummer Jr WD. 1998. Power and sample size calculations for studies involving linear regression. Controlled Clinical Trials. 19:589-601.

Eller AR, Delson E, Guthrie EH, Frost SR. 2014. Measurement protocol considerations for the cercopithecid appendicular skeleton: Body mass and function. American Journal of Physical Anthropology. 153: 113-113.

Elliott M, Collard M. 2009. Fordisc and the determination of ancestry from cranial measurements. Biology Letters. 5(6):849-852.

Elliott M, Kurki H, Weston DA, Collard M. 2014. Estimating fossil hominin body mass from cranial variables: an assessment using CT data from modern humans of known body mass. American Journal of Physical Anthropology. 154:201-214.

Elliott M, Kurki H, Weston DA, and Collard M. (submitted). Estimating body mass from post-cranial variables: an evaluation of current equations using a large known-mass sample of modern humans. American Journal of Physical Anthropology.

Ericksen MF. 1982. How “representative” is the Terry Collection? Evidence from the proximal femur. American Journal of Physical Anthropology. 59:345-350.

Feldesman MR, Fountain RL. 1996. “Race” specificity and the femur/stature ratio. American Journal of Physical Anthropology. 100(2):207-224.

Felsenstein J. 1985. Phylogenies and the comparative method. American Naturalist. 125:1–15.

Fleagle JG. 1978. Size Distributions of Living and Fossil Primate Faunas. Paleobiology. 4(1):67-76.

Frayer D. 1984. Biological and cultural change in the European Late Pleistocene and Early Holocene. In: Smith F, Spencer F, editors. The Origins of Modern Humans: A World Survey of the Fossil Evidence. New York: Wiley-Liss. p 1-50.

Fully G. 1956. Une nouvelle méthode de détermination de la taille. Annales de Medecine Legale. 35:266–273.

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

Gingerich PD. 1977. Correlation of tooth size and body size in living hominoid primates, with a note on relative brain size in Aegyptopithecus and Proconsul. American Journal of Physical Anthropology. 47(3):395-398.

Grine FE, Jungers WL, Tobias PV, Pearson OM. 1995. Fossil Homo femur from Berg Aukas, northern Namibia. American Journal of Physical Anthropology. 97(2):151-185.

Gunz P, Mitteroecker P, Neubauer S, Weber GW, Bookstein FL. 2009. Principles for the virtual reconstruction of hominin crania. Journal of Human Evolution. 57(1):48-62.

Haglund WD, Sorg MH, editors. 2002. Advances in Forensic Taphonomy: Method, Theory, and Archaeological Perspectives. Boca Raton: CRC Publishing. 544 p.

Harrell FE, Lee KL, Mark DB. 1996. Tutorial in biostatistics multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine. 15:361-387.

Hartwig-Scherer S. 1993. Body weight prediction in early fossil hominids: Towards a taxon-“independent” approach. American Journal of Physical Anthropology. 92(1):17-36.

Hartwig-Scherer S, Martin RD. 1992. Allometry and prediction in hominoids: A solution to the problem of intervening variables. American Journal of Physical Anthropology. 88:37-57.

Harvey P, Martin R, Clutton-Brock T. 1987. Life histories in comparative perspective. In: Smuts B, Cheney D, Seyfarth R, Wrangham W, Struhsaker T, editors. Primate Societies. Stuttgart: Thieme-Verlag. p 41-34.

Henderson AM, Corruccini RS. 1976. Relationship between tooth size and body size in American Blacks. Journal of Dental Research. 55(1):94-96.

Henneberg M, Stephan CN, Norris RM. 2005. Sources of biological variation. Is sex really important? American Journal of Physical Anthropology. 126(S40):114.

Hennig C, Cooper D. 2011. Brief communication: the relation between standard error of the estimate and sample size of histomorphometric aging methods. American Journal of Physical Anthropology. 145(4):658-664.

Hens SM, Konigsberg LW, Jungers WL. 2000. Estimating stature in fossil hominids: which regression model and reference sample to use? Journal of Human Evolution. 38(6):767-784.

Hinton P. 2004. Statistics explained. 2nd ed. New York: Routledge. 379 p.

Holliday TW. 2002. Body size and postcranial robusticity of European Upper Paleolithic hominins. Journal of Human Evolution. 43(4):513-528.

Holloway RL. 1980. Within-species brain-body weight variability: A reexamination of the Danish data and other primate species. American Journal of Physical Anthropology. 53(1):109-121.

Hounsfield G. 1973. Computerized transverse axial scanning (tomography): Part I. Description of system. British Journal of Radiology. 46:1016-1022.

Hounsfield G. 1980. Computed medical imaging. Science. 210(4465):22-28.

Hruschka DJ, Rush EC, Brewis AA. 2013. Population differences in the relationship between height, weight, and adiposity: An application of Burton's model. American Journal of Physical Anthropology. 151(1):68-76.

Humphrey LT. 1998. Growth patterns in the modern human skeleton. American Journal of Physical Anthropology. 105(1):57-72.

Hylander WL. 1985. Mandibular Function and Biomechanical Stress and Scaling. American Zoologist. 25(2):315-330.

International Union for Conservation of Nature. 2014. CITES initiatives and news. http://www.iucn.org/news_homepage/events/cities/

Janssen I, Heymsfield SB, Wang Z, Ross R. 2000. Skeletal muscle mass and distribution in 468 men and women aged 18–88 yr. Journal of Applied Physiology. 89(1):81-88.

Jantz RL, Ousley SD. 2005. FORDISC 3.0: Personal computer forensic discriminant functions. Knoxville: University of Tennessee.

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. Chinese Science Bulletin 1-9.

Jung, GU, Lee, UY, Kim, DH, Kwak, DS, Ahn, YW, Kim, YS. 2014. The body mass estimation from human talus: inductive approach and 3D morphometric study. The FASEB Journal. 28(S1):919-1.

Jungers WL. 1988. Relative joint size and hominoid locomotor adaptations with implications for the evolution of hominid bipedalism. Journal of Human Evolution. 17:247-265.

Jungers WL. 1990a. Problems and methods in reconstructing body size in fossil primates. In: Damuth J, MacFadden BJ, editors. Body Size in Mammalian Paleobiology. Cambridge: Cambridge University Press. p. 103–118.

Jungers WL. 1990b. Scaling of hominoid femoral head size and the evolution of hominid bipedalism. American Journal of Physical Anthropology. 81(2):246.

Jungers WL, Minns RJ. 1979. Computed tomography and biomechanical analysis of fossil long bones. American Journal of Physical Anthropology. 50(2):285-290.

Kappelman J. 1996. The evolution of body mass and relative brain size in fossil hominids. Journal of Human Evolution. 30(3):243-276.

Kettner M, Schmidt P, Potente S, Ramsthaler F, Schrodt M. 2011. Reverse Engineering - Rapid Prototyping of the Skull in Forensic Trauma Analysis. Journal of Forensic Sciences. 56(4):1015-1017.

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

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

Knüsel C, Batt C, Cook G, Montgomery J, Müldner G, Ogden A, Palmer C, Stern B, Todd J, Wilson A. 2010. The Identity of the St Bees Lady, Cumbria: An Osteobiographical Approach. Medieval Archaeology. 54(1):271-311.

Komar DA, Grivas CR. 2008. Manufactured populations: what do contemporary reference skeletal collections represent? A comparative study using the Maxwell Museum documented collection. American Journal of Physical Anthropology. 137(2):224-233.

Konigsberg LW, Frankenberg SR. 2013. Bayes in biological anthropology. American Journal of Physical Anthropology. 152(S57):153-184.

Konigsberg LW, Hens SM, Jantz LM, Jungers WL. 1998. Stature estimation and calibration: Bayesian and maximum likelihood perspectives in physical anthropology. American Journal of Physical Anthropology. 107(S27):65-92.

Kordos L, Begun DR. 2001. Primates from Rudabanya: allocation of specimens to individuals, sex and age categories. Journal of Human Evolution. 40(1):17-39.

Kuczmarski RJ, Flegal KM. 2000. Criteria for definition of overweight in transition: background and recommendations for the United States. The American Journal of Clinical Nutrition 72(5):1074-1081.

Kurki HK, Ginter JK, Stock JT, Pfeiffer S. 2010. Body size estimation of small-bodied humans: Applicability of current methods. American Journal of Physical Anthropology. 141(2):169-180.

Lam YM, Pearson OM. 2005. Bone density studies and the interpretation of the faunal record. Evolutionary Anthropology: Issues, News, and Reviews 14(3):99-108.

Lauterbur PC. 1973. Image formation by induced local interactions: examples employing nuclear magnetic resonance. Nature. 242(5394):190-191.

Leakey L. 1959. A new fossil skull from Olduvai. Nature. 184:4685.

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

Lewin P. 1988. First stereoscopic images from CT reconstructions of mummies. American Journal of Roentgenology 151(6):1249.

Lieberman DE, Devlin MJ, Pearson OM. 2001. Articular area responses to mechanical loading: effects of exercise, age, and skeletal location. American Journal of Physical Anthropology. 116(4):266-277.

Lieberman DE, Polk JD, Demes B. 2004. Predicting long bone loading from cross-sectional geometry. American Journal of Physical Anthropology. 123(2):156-171.

Lopes PML, Moreira CR, Perrella A, Antunes JL, Cavalcanti, MGP. 2008. 3-D volume rendering maxillofacial analysis of angular measurements by multislice CT. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology. 105:224-230.

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 Science International. 231(1):405-e1.

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

Martin RA. 1981. On extinct hominid population densities. Journal of Human Evolution. 10(5):427-428.

McHenry HM. 1976. Early hominid body weight and encephalization. American Journal of Physical Anthropology. 45(1):77-83.

McHenry HM. 1988. New estimates of body weight in early hominids and their significance to encephalization and megadontia in “robust” australopithecines. In: Grine FE, editor. Evolutionary History of the" robust" Australopithecines. New Jersey: Transaction Publishers. p. 327-337.

McHenry HM. 1992. Body size and proportions in early hominids. American Journal of Physical Anthropology. 87(4):407-431.

McHenry HM, Coffing K. 2000. Australopithecus to Homo: Transformations in body and mind. Annual Review of Anthropology. 29:125-14.

Melton N, Montgomery J, Knusel C, Batt C, Needham S, Pearson MP, Sheridan A, Heron C, Horsley T, Schmidt A. 2010. Gristhorpe Man: an Early Bronze Age log-coffin burial scientifically defined. Antiquity. 84(325):796-815.

Messmer P, Matthews F, Jacob A, Kikinis R, Regazzoni P, Noser H. 2007. A CT Database for Research, Development and Education: Concept and Potential. Journal of Digital Imaging. 20(1):17-22.

Moore MK, Schaefer E. 2011. A Comprehensive Regression Tree to Estimate Body Weight from the Skeleton. Journal of Forensic Sciences. 56(5):8.

Myszka A, Piontek J, Vancata A. 2012. Body mass reconstruction on the basis of selected skeletal traits. Anthropologischer Anzeige. 69:305-315.

Nakatsukasa M, Pickford M, Egi N, Senut B. 2007. Femur length, body mass, and stature estimates of Orrorin tugenensis, a 6 Ma hominid from Kenya. Primates. 48(3):171-178.

Niskanen M, Junno J-A. 2009. Estimation of African apes’ body size from postcranial dimensions. Primates. 50(3):211-220.

Oleksiak DA. 1986. The Estimation Of Body-Weight For Neandertals And Early Anatomically Modern Humans. American Journal of Physical Anthropology. 69(2):248-248.

Olivarius NF, Andreasen AH, Loken J. 1997. Accuracy of 1-, 5- and 10-year body weight recall given in a standard questionnaire. International Journal of Obesity and related Metabolic Disorders. 21:67-71.

Pearson K. 1899. Mathematical Contributions to the Theory of Evolution. V. On the Reconstruction of the Stature of Prehistoric Races. Philosophical Transactions of the Royal Society of London Series A. 169-244.

Pearson OM, Lieberman DE. 2004. The aging of Wolff's “law”: ontogeny and responses to mechanical loading in cortical bone. American Journal of Physical Anthropology. 125(S39):63-99.

Pearson O, Cordero R, Busby A. 2008. How different were Neanderthals' habitual activities? A comparative analysis with diverse groups of recent humans. In: Hublin J-J, Harvati K, Harrison T, editors. Neanderthals Revisited: New Approaches and Perspectives. Dordrecht: Springer. p. 135-156.

Perissinotto E, Pisent C, Sergi G, Grigoletto F, Enzi G, Group IW. 2002. Anthropometric measurements in the elderly: age and gender differences. British Journal of Nutrition. 87(2):177-186.

Perry GS, Byers TE, Mokdad AH, Serdula MK, Williamson DF. 1995. The validity of self-reports of past body weights by U.S. adults. Epidemiology. 6:61-66.

Plavcan JM. 2012. Body size, size variation, and sexual size dimorphism in early Homo. Current Anthropology. 53:S409-S423.

Pokines J, Symes SA, Roper C. 2013. Manual of forensic taphonomy. Boca Raton: CRC Press. 496 p.

Pomeroy E, Stock JT. 2012. Estimation of stature and body mass from the skeleton among coastal and mid-altitude andean populations. American Journal of Physical Anthropology. 147(2):264-279.

Porter AMW. 1999. The prediction of physique from the skeleton. International Journal of Osteoarchaeology. 9(2):102-115.

Pothiwala P, Evans EM, Chapman-Novakofski KM. 2006. Ethnic Variation in Risk for Osteoporosis among Women: A Review of Biological and Behavioral Factors. Journal of Women's Health. 15(6):709-719.

Power ML, Schulkin J. 2008. Sex differences in fat storage, fat metabolism, and the health risks from obesity: possible evolutionary origins. British Journal of Nutrition. 99:931-940.

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

Rainwater C, Cabo-Perez L, Symes S. 2007. Body mass estimation and personal identification. American Journal of Physical Anthropology. 132(S44): 194-195.

Raxter MH, Auerbach BM, Ruff CB. 2006. Revision of the Fully technique for estimating statures. American Journal of Physical Anthropology. 130(3):374-384.

Reichs K, Dorion R. 1992. Use of Computed Tomography (CT) Scans in the Comparison of Frontal Sinus Configurations. Canadian Society of Forensic Science. 25(1).

Rightmire GP. 1986. Body size and encephalization in Homo erectus Anthropos. 23:139-149.

Rightmire GP. 2004. Brain size and encephalization in early to mid-Pleistocene Homo. American Journal of Physical Anthropology. 124(2):109-123.

Robinson C, Eisma R, Morgan B, Jeffery A, Graham EAM, Black S, Rutty GN. 2008. Anthropological Measurement of Lower Limb and Foot Bones Using Multi-Detector Computed Tomography. Journal of Forensic Sciences. 53(6):1289-1295.

Robson SL, Wood B. 2008. Hominin life history: reconstruction and evolution. Journal of Anatomy. 212(4):394-425.

Rosenberg KR, Zuné L, Ruff CB. 2006. Body size, body proportions, and encephalization in a Middle Pleistocene archaic human from northern China. Proceedings of the National Academy of Sciences USA. 103(10):3552-3556.

Ross CF. 1995. Allometric and functional influences on primate orbit orientation and the origins of the Anthropoidea. Journal of Human Evolution. 29:201-227.

Ruff CB. 1987. Structural allometry of the femur and tibia in hominoidea and macaca. Folia Primatologica. 48(1).

Ruff CB. 1991. Climate and body shape in hominid evolution. Journal of Human Evolution. 21(2):81-105.

Ruff CB. 1994. Morphological adaptation to climate in modern and fossil hominids. American Journal of Physical Anthropology. 37(S19):65-107.

Ruff CB. 2000a. Body mass prediction from skeletal frame size in elite athletes. American Journal of Physical Anthropology. 113(4):507-517.

Ruff CB. 2000b. Body size, body shape, and long bone strength in modern humans. Journal of Human Evolution. 38(2):269-290.

Ruff CB. 2002. Variation in Human Body Size and Shape. Annual Review of Anthropology. 31:211-232.

Ruff CB. 2003. Long bone articular and diaphyseal structure in Old World monkeys and apes. II: estimation of body mass. American Journal of Physical Anthropology. 120(1):16-37.

Ruff CB. 2010. Body size and body shape in early hominins-implications of the Gona pelvis. Journal of Human Evolution. 58(2):166.

Ruff CB, Hayes W. 1983. Cross-sectional geometry of Pecos Pueblo femora and tibiae—A biomechanical investigation: I. Method and general patterns of variation. American Journal of Physical Anthropology. 60(3):359-381.

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

Ruff CB, Holt B, Trinkaus E. 2006a. Who's afraid of the big bad Wolff?: “Wolff's law” and bone functional adaptation. American Journal of Physical Anthropology. 129(4):484-498.

Ruff CB, Holt BM, Sládek V, Berner M, Murphy Jr WA, zur Nedden D, Seidler H, Recheis W. 2006b. Body size, body proportions, and mobility in the Tyrolean “Iceman”. Journal of Human Evolution. 51(1):91-101.

Ruff CB, Holt BM, Niskanen M, Sladék V, Berner M, Garofalo E, Garvin HM, Hora M, Maijanen H, Niinimäki S et al. 2012. Stature and body mass estimation from skeletal remains in the European Holocene. American Journal of Physical Anthropology. 148(4):601-617.

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

Ruff CB, Scott WW, Liu AYC. 1991. Articular and diaphyseal remodeling of the proximal femur with changes in body mass in adults. American Journal of Physical Anthropology. 86(3):397-413.

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

Ruff CB, Trinkaus E, Walker A, Larsen CS. 1993. Postcranial robusticity in Homo. I: Temporal trends and mechanical interpretation. American Journal of Physical Anthropology. 91(1):21-53.

Rühli FJ, Lanz C, Ulrich-Bochsler S, Alt KW. 2002. State-of–the–art imaging in palaeopathology: the value of multislice computed tomography in visualizing doubtful cranial lesions. International Journal of Osteoarchaeology. 12(5):372-379.

Scientific Working Group for Forensic Anthropology. 2012. Stature Estimation. Revision 1. Issued: 08/02/2012. http://swganth.startlogic.com/Stature%20Estimation%20Rev%201.pdf.

Sciulli PW, Blatt SH. 2008. Evaluation of juvenile stature and body mass prediction. American Journal of Physical Anthropology. 136(4):387-393.

Seidell JC, Visscher TL. 2000. Body weight and weight change and their health implications for the elderly. European Journal of Clinical Nutrition. 54(S3):S33-39.

Seidler H, Bernhard W, Teschler-Nicola M, Platzer W, zur Nedden D, Henn R, Oberhauser A, Sjovold T. 1992. Some anthropological aspects of the prehistoric Tyrolean ice man. Science 258:455-457.

Shen W, Punyanitya M, Wang ZM, Gallagher D, St. Onge, MP. 2004. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. Journal of Applied Physiology. 97:2333–2338.

Smith RJ. 1996. Biology and Body Size in Human Evolution: Statistical Inference Misapplied. Current Anthropology. 37(3):451-481.

Smith RJ. 2002. Estimation of Body Mass in Paleontology. Journal of Human Evolution. 43(2):271-287.

Smith RJ. 2009. Use and misuse of the reduced major axis for line-fitting. American Journal of Physical Anthropology. 140:476-486.

Smith RJ, Jungers WL. 1997. Body mass in comparative primatology. Journal of Human Evolution. 32(6):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 Proceedings of the American Academy of Forensic Sciences. 28:361.

Snyder WS, Cook MJ, Nasset ES. 1975. Report of the Task Group on Reference Man. Oxford: Pergamon. 480 p.

Sokal RR, Rohlf FJ. 2012. Biometry (4th ed). New York: WH Freeman & Co. 937 p.

Spahic D, Karac A. 2009. Using Magnetic Resonance Images To Create 3D Models Of Bones For Subsequent Numerical Analysis. 13th International Research/Expert Conference: ”Trends in the Development of Machinery and Associated Technology”. Hammamet, Tunisia.

Spocter MA, Manger PR. 2007. The use of cranial variables for the estimation of body mass in fossil hominins. American Journal of Physical Anthropology. 134(1):92-105.

Spoor CF, Zonneveld FW, Macho GA. 1993. Linear measurements of cortical bone and dental enamel by computed tomography: Applications and problems. American Journal of Physical Anthropology. 91(4):469-484.

Steckel RH, Rose JCT. 2002. The Backbone of History: Health and Nutrition in the Western Hemisphere. New York: Cambridge University Press.654 p.

Steudel K. 1980. New estimates of early hominid body size. American Journal of Physical Anthropology. 52(1):63-70.

Steudel K. 1985. Allometric perspectives on fossil catarrhine morphology. In: Size and scaling in primate biology. New York: Springer. p. 449-475.

Stirland AJ. 1998. Musculoskeletal evidence for activity: problems of evaluation. International Journal of Osteoarchaeology. 8(5):354-362.

Stubblefield PR. 2002. Consideration of fatness in body mass estimation from skeletal indicators. American Journal of Physical Anthropology. 117(S34):11-34.

Suskewicz JA. 2004. Estimation Of Living Body Weight Based On Measurements Of Anterior Superior Iliac Spine Breadth And Stature. Unpublished Masters thesis. Baton Rouge: Louisiana State University.

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

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

Tate JR, Cann CE. 1982. High-resolution computed tomography for the comparative study of fossil and extant bone. American Journal of Physical Anthropology. 58(1):67-73.

Thali MJ, Jackowski C, Oesterhelweg L, Ross SG, Dirnhofer R. 2007. VIRTOPSY - The Swiss virtual autopsy approach. Legal Medicine. 9(2):100-104.

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

Tobias PV. 2001. Re-creating ancient hominid virtual endocasts by CT-scanning. Clinical Anatomy. 14(2):134-141.

Trinkaus E, Buzhilova AP, Mednikova MB, Dobrovolskaya MV. 2014. The People of Sunghir: Burials, Bodies, and Behavior in the Earlier Upper Paleolithic. Oxford: Oxford University Press. 368 p.

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

Trinkaus E, Churchill SE, Ruff CB. 1994. Postcranial robusticity in Homo. II: Humeral bilateral asymmetry and bone plasticity. American Journal of Physical Anthropology. 93(1):1-34.

Trinkaus E, Jelínek J. 1997. Human remains from the Moravian Gravettian: the Dolní Věstonice 3 postcrania. Journal of Human Evolution. 33(1):33-82.

Trinkaus E, Ruff, CB. 1999. Diaphyseal cross-sectional geometry of Near Eastern Middle Paleolithic humans: the femur. Journal of Archaeological Science. 26:409–424.

Trinkaus E, Ruff CB. 2012. Femoral and tibial diaphyseal crosssectional geometry in Pleistocene Homo. Paleoanthropology. 13:62.

Trotter M, Gleser GC. 1952. Estimation of stature from long bones of American Whites and Negroes. American Journal of Physical Anthropology. 10(4):463-514.

Uhl NM, Rainwater CW, Konigsberg LW. 2013. Testing for size and allometric differences in fossil hominin body mass estimation. American Journal of Physical Anthropology. 151(2):215-229.

Vannucci RC, Barron TF, Lerro D, Antón SC, Vannucci SJ. 2011. Craniometric measures during development using MRI. NeuroImage. 56(4):1855-1864.

Vercellotti G, Stout SD, Boano R, Sciulli PW. 2011. Intrapopulation variation in stature and body proportions: Social status and sex differences in an Italian medieval population (Trino Vercellese, VC). American Journal of Physical Anthropology. 145(2), 203-214.

Walker R, Gurven M, Hill K, Migliano A, Chagnon N, De Souza R, Djurovic G, Hames R, Hurtado AM, Kaplan H et al. 2006. Growth rates and life histories in twenty-two small-scale societies. American Journal of Human Biology. 18(3):295-311.

Walker MJ, Ortega J, Parmovd K, Lopez MV, Trinkaus E. 2011. Morphology, body proportions, and postcranial hypertrophy of a female Neandertal from the Sima de las Palomas, southeastern Spain. Proceedings of the National Academy of Sciences USA. 108:10087-10091.

Weiss E, Corona L, Schultz B. 2012. Sex differences in musculoskeletal stress markers: problems with activity pattern reconstructions. International Journal of Osteoarchaeology. 22(1):70-80.

Whitmarsh T, Humbert L, De Craene M, Barquero LMdR, Fritscher K, Schubert R, Eckstein F, Link T, Frangi AF. 2010. 3D bone mineral density distribution and shape reconstruction of the proximal femur from a single simulated DXA image: an in vitro study. In: Dawant BM, Haynor DR, editors. San Diego: SPIE Medical Imaging. International Society for Optics and Photonics. pp. 76234U-76234U.

Wind J. 1984. Computerized X-ray tomography of fossil hominid skulls. American Journal of Physical Anthropology. 63(3):265-282.

Wood BA. 1979. An analysis of tooth and body size relationships in five primate taxa. Folia primatologica. 31(3):187-211.

Wood BA. 2011. Wiley-Blackwell encyclopedia of human evolution. New York: Wiley-Blackwell. 1056 p.

Wood BA, Collard M. 1996. Comments and Reply to Smith: Biology and Body Size in Human Evolution: Statistical Inference Misapplied. Current Anthropology. 37(3):451-481.

Wood BA, Collard M. 1999. The human genus. Science. 284:65-71.

Woodhead HJ, Kemp AF, Blimkie CJR, Briody JN, Duncan CS, Thompson M, Lam A, Howman-Giles R, Cowell CT. 2001. Measurement of Midfemoral Shaft Geometry: Repeatability and Accuracy Using Magnetic Resonance Imaging and Dual-Energy X-ray Absorptiometry. Journal of Bone and Mineral Research. 16(12):2251-2259.

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. Pharmacological Research. 32:241-248.

Zollikofer CPE, Ponce de Leon MS. 2002. Visualizing patterns of craniofacial shape variation in Homo sapiens. Proceedings of the Royal Society of London Series B: Biological Sciences. 269(1493):801-807.

Zollikofer CPE, Ponce de Leon MS, Lieberman DE, Guy F, Pilbeam D, Likius A, Mackaye HT, Vignaud P, Brunet M. 2005. Virtual cranial reconstruction of Sahelanthropus tchadensis. Nature. 434(7034):755-759.

Appendix.

Supplementary Tables

Table A1. Summary data for cranial variables (see Table 2.2 for descriptions)

	Female (n=125)		Male (n=128)		Combined-sex (n=253)	
Variable	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
BORB	35.3 (1.7)	31.9-39.7	37.1 (1.8)	33.7-42.3	36.2 (1.9)	31.9-42.3
HORB	34.1 (2.2)	28.3-39.9	34.4 (2.2)	27.5-41.4	34.3 (2.2)	27.5-41.4
BIOR	93.7 (3.8)	84.7-105.9	98.3 (3.9)	88.3-107.1	96.0 (4.5)	84.7-107.1
BPOR	112.2 (5.0)	99.5-123.3	119.6 (5.1)	109.6-132.3	116.0 (6.3)	99.5-132.3
LFM	33.4 (2.2)	29.4-41.2	35.0 (2.7)	30.0-42.1	34.2 (2.5)	29.4-42.1
BFM	28.4 (2.2)	22.9-34.1	29.7 (2.5)	22.1-37.3	29.1 (2.4)	22.1-37.3
ORBA1	1204.0 (104.1)	911.2-1501.5	1274.7 (103.3)	1035.0-1531.9	1239.8 (109.4)	911.3-1531.9
ORBA2	945.6 (81.8)	715.7-1179.3	1001.2 (81.1)	812.9-1203.1	973.7 (85.9)	715.7-1203.1
ORBA3	967.3 (82.4)	732.8-1205.7	1021.4 (81.0)	831.7-1267.9	994.3 (85.9)	732.8-1267.9
FMA1	951.0 (118.8)	692.3-1404.9	1044.2 (148.5)	758.0-1507.2	998.1 (142.4)	692.3-1507.2
FMA2	746.9 (93.3)	543.7-1103.4	820.1 (116.6)	595.4-1183.7	783.9 111.7	543.7-1183.7
FMA3	671.3 (86.7)	490.5-1008.8	742.4 (102.4)	548.1-1059.1	706.9 101.2	490.5-1059.1

Table A2. Difference between known and estimated mass, Aiello and Wood (1994) RMA regressions for cranial variables

	Female (n=125)			Male (n=128)			Combined-sex (n=253)		
Variable	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)
BORB	11.6 (32.8)	27.9 (20.7)*	43.2	3.6 (31.4)*	25.8 (18.1)	39.8	7.6 (32.3)*	26.8 (19.4)	41.5
HORB	28.5 (28.6)	35.6 (19.0)*	24.0	38.4 (22.6)*	40.8 (17.8)	17.2	33.5 (26.1)*	38.2 (18.6)	20.6
BIOR	-11.3 (35.1)	25.7 (26.3)	56.8	-17.8 (32.0)*	26.1 (25.6)	51.6	-14.6 (33.7)*	25.9 (25.9)	54.2
BPOR	8.4 (26.9)	22.0 (17.6)*	53.6	3.5 (20.8)*	17.2 (12.2)	59.4	5.9 (24.1)*	19.5 (15.3)	56.5
LFM	-24.5 (41.9)	34.8 (33.7)*	43.2	-21.3 (37.6)*	31.1 (30.0)	46.9	-22.9 (39.8)*	32.9 (31.9)	45.1
BFM	-63.0 (66.8)	69.6 (59.8)*	22.4	-62.1 (61.7)*	66.0 (57.4)	16.4	-62.5 (64.1)*	67.8 (58.5)	19.4
ORBA1	19.7 (26.9)	27.7 (18.4)*	36.0	23.1 (21.8)*	27.5 (15.8)	35.2	21.4 (24.4)*	27.6 (17.1)	35.6
FMA1	-41.6 (49.7)	47.4 (44.2)*	35.2	-39.1 (43.0)*	43.4 (38.7)	32.8	-40.4 (46.4)*	45.4 (41.5)	34.0

PPE: percent prediction error (known - estimated)/known * 100, |PPE|: absolute percent prediction error, 20%: percent of individuals whose estimated body masses fall within +/-20% of known mass. 1. Directional differences (positive values indicate underestimation, negative values indicate overestimation); 2. Absolute differences; * indicates significance at p=0.01.

Table A3. Difference between known and estimated mass, Spocter and Manger (2007) RMA regressions for cranial variables

	Female (n=125)			Male (n=128)			Combined-sex (n=253)		
Variable	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)
BORB	43.6 (18.3)*	44.5 (16.0)	8.8	42.6 (15.7)	42.8 (15.4)	9.4	43.1 (17.0)*	43.6 (15.7)	9.1
HORB	76.2 (9.5)*	76.2 (9.5)	0.0	79.5 (7.5)	79.5 (7.5)	0.0	77.8 (8.7)*	77.8 (8.7)	0.0
BIOR	74.4 (7.3)*	74.4 (7.3)	0.0	74.4 (6.2)	74.4 (6.2)	0.0	74.4 (6.8)*	74.4 (6.8)	0.0
BPOR	-47.3 (47.2)*	52.2 (41.7)	20.8	-65.8 (39.6)	66.0 (39.3)	11.7	-56.7 (44.4)*	59.2 (41.0)	16.2
LFM	32.2 (24.8)*	36.9 (17.0)	16.8	32.1 (23.5)	35.3 (18.4)	23.4	32.2 (24.1)*	36.1 (17.7)	20.2
BFM	6.9 (37.2)*	30.4 (22.3)	38.4	8.2 (33.6)	26.8 (21.8)	46.1	7.5 (35.4)*	28.6 (22.1)	42.3
ORBA1	64.8 (11.2)*	64.8 (11.2)	0.0	66.9 (8.8)	66.9 (8.8)	0.0	65.9 (10.1)*	65.9 (10.1)	0.0
ORBA2	-22302.8 (9933.2)*	22302.8 (9933.2)	0.0	-23737.0 (9749.3)	23737.0 (9749.3)	0.0	23028.4 (9847.2)*	23028.4 (9847.2)	0.0
FMA1	22.4 (28.0)*	31.13 (17.6)	34.4	23.0 (24.8)	29.7 (16.2)	25.8	22.7 (26.4)*	30.4 (16.9)	30.0
FMA2	-15108.0 (9350.4)*	15108.0 (9350.4)	0.0	-18210.7 (11742.3)	18210.7 (11742.3)	0.0	-16677.7 (10720.3)*	16677.7 (10720.3)	0.0

PPE: percent prediction error (known - estimated)/known * 100, |PPE|: absolute percent prediction error, 20%: percent of individuals whose estimated body masses fall within +/-20% of known mass. 1. Directional differences (positive values indicate underestimation, negative values indicate overestimation); 2. Absolute differences; * indicates significance at p=0.01.

Table A4. Difference between known and estimated masses using individual data and mean data for 50 sub-samples of ten individuals, Aiello and Wood (1994) LS regressions for cranial variables

	Individual data	50 x 10 individuals	Significant Difference[2]
Variable	**Mean PPE[1] (SD)**	**Mean PPE[1] (SD)**	**p=0.01**
BORB	8.0 (31.3)	15.8 (9.1)	Higher
HORB	31.8 (26.4)	37.0 (6.8)	Higher
BIOR	-12.9 (32.3)	-6.0 (11.5)	Lower
BPOR	0.9 (25.2)	6.0 (8.4)	Not sign. different
LFM	-21.0 (37.1)	-14.1 (10.8)	Not sign. different
BFM	-54.0 (57.5)	-42.5 (19.7)	Not sign. different
ORBA1	23.9 (23.4)	28.9 (6.4)	higher
FMA1	-32.9 (41.8)	-25.9 (13.7)	Not sign. different

1. Directional differences (positive values indicate underestimation, negative values indicate overestimation). 2. indicates whether the bootstrapped 50 x 10-individual mean PPE is higher, lower or not significantly different from the single-individual sample mean PPE.

Table A5. Summary data for the BMI-Restricted sample (BMI=18.5-25)

	Female (n=57)		Male (n=59)		Combined-sex (n=116)	
Variable	**Mean (SD)**	**Range**	**Mean (SD)**	**Range**	**Mean (SD)**	**Range**
Weight (kg)	59.4 (8.1)	43.0-95.0	72.3 (8.4)	52.0-86.1	65.9 (10.4)	43.0-95.0
Stature (cm)	165.9 (8.2)	150.0-195.0	177.9 (8.6)	156.0-193.0	172.0 (10.4)	150.0-195.0
Age (yrs)	49.7 (16.2)	20.0-89.0	45.6 (13.9)	22.0-80.0	47.6 (15.2)	20.0-89.0
BMI	21.6 (2.0)	18.6-24.9	22.8 (1.5)	19.5-24.9	22.2 (1.9)	18.6-24.9

Table A6. Summary data for the cranial variables, BMI-Restricted sample

	Female (n=57)		Male (n=59)		Combined-sex (n=116)	
Variable	**Mean (SD)**	**Range**	**Mean (SD)**	**Range**	**Mean (SD)**	**Range**
BORB	35.5 (1.6)	32.0-39.7	37.2 (1.8)	33.7-41.2	36.3 (1.9)	32.0-41.2
HORB	34.0 (2.1)	28.6-39.9	34.3 (2.3)	28.5-41.4	34.2 (2.2)	28.5-41.4
BIOR	94.0 (3.8)	87.4-105.9	97.9 (4.2)	88.3-107.1	96.0 (4.4)	87.4-107.1
BPOR	111.7 (4.7)	103.1-122.7	119.0 (4.8)	110.6-132.3	115.4 (6.0)	103.1-132.2
LFM	33.4 (2.4)	29.6-41.2	35.1 (2.7)	30.1-42.1	34.3 (2.7)	29.6-42.1
BFM	28.5 (2.1)	22.9-34.1	30.0 (2.4)	26.1-37.3	29.2 (2.4)	22.9-37.3
ORBA1	1205.1 (96.9)	915.2-1420.4	1275.0 (97.5)	1035.0-1519.4	1240.6 (103.0)	915.2-1519.4
ORBA2	946.4 (76.1)	718.8-1115.6	1001.4 (76.6)	812.9-1193.3	974.4 (80.9)	718.8-1193.3
ORBA3	966.0 (75.8)	732.8-1118.3	1014.5 (78.1)	831.7-1205.2	990.6 (80.5)	732.8-1205.2
FMA1	953.5 (121.4)	744.3-1404.9	1055.0 (147.7)	802.2-1507.2	1005.1 (144.2)	744.3-1507.2
FMA2	748.9 (95.3)	584.5-1103.4	828.6 (116.0)	630.0-1183.7	789.4 (113.2)	584.5-1183.7
FMA3	673.1 (85.3)	530.7-1008.8	747.3 (100.3)	554.2-1040.2	710.8 (100.0)	530.7-1040.2

Table A7. Difference between known and estimated mass, Aiello and Wood (1994) LS regressions for cranial variables, BMI-restricted samples

	Female (n=57)			Male (n=59)			Combined-sex (n=116)		
Variable	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)
BORB	-0.8 (26.6)	19.8 (17.6)	63.2	-6.0 (28.7)	22.9 (18.0)	50.8	-3.4 (27.7)	21.4 (17.8)	56.9
HORB	19.9 (24.5)*	26.4 (17.2)	43.9	30.5 (23.8)*	34.6 (17.2)	27.1	25.3 (24.6)*	30.6 (17.6)	35.3
BIOR	-24.6 (25.4)*	26.7 (23.1)	50.9	-25.1 (28.9)*	28.7 (25.3)	49.1	-24.9 (27.1)*	27.7 (24.2)	50.0
BPOR	-5.2 (18.8)	14.7 (12.7)	71.9	-9.6 (17.6)*	15.6 (12.6)	72.9	-7.5 (18.3)*	15.1 (12.6)	72.4
LFM	-36.6 (29.1)*	37.2 (28.4)	28.1	-31.0 (31.3)*	33.7 (28.3)	42.3	-33.8 (30.3)*	35.4 (28.3)	35.3
BFM	-71.5 (41.1)*	73.0 (38.4)	8.8	-71.9 (53.0)*	71.9 (53.0)	6.7	-71.7 (47.3)*	72.4 (46.2)	7.8
ORBA1	13.3 (19.5)*	19.2 (13.6)	59.6	18.2 (18.6)*	22.4 (13.1)	42.3	15.8 (19.1)*	20.8 (13.4)	50.9
FMA1	-49.1 (28.5)*	49.2 (28.4)	15.8	-46.0 (34.5)*	46.0 (34.5)	17.0	-47.5 (31.6)*	47.6 (31.5)	16.4

PPE: percent prediction error (known - estimated)/known * 100, |PPE|: absolute percent prediction error, 20%: percent of individuals whose estimated body masses fall within +/-20% of known mass. 1. Directional differences (positive values indicate underestimation, negative values indicate overestimation); 2. Absolute differences; * indicates significance at p=0.01.

Table A8. Difference between known and estimated mass, Kappelman (1996) LS regressions for cranial variables, BMI-restricted samples

	Female (n=57)			Male (n=59)			Combined-sex (n=116)		
Variable	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)
HORB	17.5 (25.3)*	25.4 (17.3)	45.6	28.4 (24.7)*	33.4 (17.2)	28.8	23.0 (25.5)*	29.4 (17.6)	37.1
ORBA3	35.3 (13.5)*	35.3 (13.5)	10.5	40.7 (12.3)*	40.7 (12.3)	8.5	38.1 (13.1)*	38.1 (13.1)	9.5

PPE: percent prediction error (known - estimated)/known * 100, |PPE|: absolute percent prediction error, 20%: percent of individuals whose estimated body masses fall within +/-20% of known mass. 1. Directional differences (positive values indicate underestimation, negative values indicate overestimation); 2. Absolute differences; * indicates significance at p=0.01.

Table A9. Difference between known and estimated mass, Spocter and Manger (2007) LS regressions for cranial variables, BMI-restricted samples

	Female (n=57)			Male (n=59)			Combined-sex (n=116)		
Variable	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)
BORB	35.8 (13.2)*	36.2 (12.2)	10.5	37.1 (13.1)*	37.2 (12.6)	11.9	36.4 (13.1)*	36.7 (12.4)	11.2
HORB	73.8 (8.1)*	73.8 (8.1)	0.0	77.2 (7.9)*	77.2 (7.9)	0.0	75.5 (8.1)*	75.5 (8.1)	0.0
BIOR	70.3 (5.1)*	70.3 (5.1)	0.0	71.4 (5.6)*	71.4 (5.6)	0.0	70.9 (5.4)*	70.9 (5.4)	0.0
BPOR	-52.5 (32.6)*	52.5 (32.6)	14.0	-69.7 (32.9)*	69.7 (32.9)	5.1	-61.3 (33.7)*	61.3 (33.7)	9.5
LFM	21.5 (21.5)*	26.8 (14.1)	31.6	21.5 (23.6)*	26.9 (16.9)	35.6	21.5 (22.5)*	26.9 (15.6)	33.6
BFM	-0.3 (24.2)	19.3 (14.4)	66.7	-0.7 (31.3)	22.7 (21.4)	55.9	-0.5 (27.9)	21.0 (18.3)	61.2
ORBA1	57.2 (8.8)*	57.2 (8.8)	0.0	60.3 (8.3)*	60.3 (8.3)	0.0	58.8 (8.6)*	58.8 (8.6)	0.0
ORBA2	-23957.4 (8299.1)*	23957.4 (8299.1)	0.0	-25136.7 (9053.1)*	25136.7 (9053.1)	0.0	-24577.2 (8673.3)*	24557.2 (8673.3)	0.0
FMA1	7.6 (20.1)*	15.90 (14.4)	70.2	7.3 (25.1)*	21.2 (15.1)	54.2	7.5 (22.7)*	18.6 (15.0)	62.1
FMA2	-15254.2 (7743.6)*	15254.2 (7743.6)	0.0	-19001.1 (11776.3)*	19001.1 (11776.3)	0.0	-17160.0 (10133.2)*	17159.9 (10133.2)	0.0

PPE: percent prediction error (known - estimated)/known * 100, |PPE|: absolute percent prediction error, 20%: percent of individuals whose estimated body masses fall within +/-20% of known mass. 1. Directional differences (positive values indicate underestimation, negative values indicate overestimation); 2. Absolute differences; * indicates significance at p=0.01.

Table A10. Mean percent prediction errors for Aiello and Wood (1994) LS regressions for cranial variables, Age-restricted samples (18-60 yrs)

	Female (n=87)		Male (n=99)		Combined-sex (n=186)	
Variable	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)
BORB	11.8 (29.4)	25.4 (18.6)	5.58 (28.7)	24.0 (16.5)	8.47 (29.1)	24.7 (17.5)
HORB	27.3 (27.5)	34.3 (18.0)	38.5 (22.9)	40.8 (18.3)	33.2 (25.7)	37.8 (18.4)
BIOR	-10.4 (29.5)	22.5 (21.6)	-14.8 (29.3)	23.1 (23.2)	-12.7 (29.4)	22.8 (22.4)
BPOR	3.03 (26.3)	20.1 (17.1)	0.11 (19.9)	16.4 (11.1)	1.48 (23.1)	18.1 (14.3)
LFM	-25.1 (36.5)	32.8 (29.7)	-18.0 (33.3)	28.1 (24.2)	-21.3 (34.9)	30.3 (27.4)
BFM	-58.2 (57.5)	63.9 (51.0)	-50.8 (53.2)	55.9 (47.7)	-54.3 (55.2)	59.7 (49.3)
ORBA1	22.7 (23.8)	28.1 (17.0)	27.1 (20.3)	29.5 (16.6)	25.0 (22.1)	28.8 (16.7)
FMA1	-37.0 (44.6)	42.0 (36.5)	-29.8 (37.1)	35.4 (31.6)	-33.2 (39.3)	38.5 (34.0)

PPE: percent prediction error (known - estimated)/known * 100. 1. Directional differences (positive values indicate underestimation, negative values indicate overestimation); 2. Absolute differences; * indicates values that are significantly different (p=0.01) from mean PPEs in age-unrestricted sample (Table 2.5).

Table A11. Percent of individuals whose mass was estimated within +/-10% or +/-15% of true mass, for postcranial equations.

Source	Female (n=125)		Male (n=128)		Combined-sex (n=253)	
	+/-10%	+/- 15%	+/-10%	+/- 15%	+/-10%	+/- 15%
FHB-1	29.6	36.0	40.6	60.2	36.0	49.8
FHB-2	29.6	43.2	39.1	54.7	34.4	49.0
FHB-3	28.0	48.0	38.3	60.9	33.2	54.6
FHB-4	31.2	42.4	39.1	57.0	34.8	49.8
Mean of FHB 1-3	28.8	45.6	40.6	60.9	34.0	53.0
Mean of FHB 1-4	28.8	48.0	40.6	59.4	33.6	54.2
STBIB-1	28.0	43.2	34.4	49.2	30.4	47.0
STBIB-2	28.0	44.0	36.7	53.1	30.8	45.5

Table A12. Differences between known and estimated mass for cranial variables (MA equations), combined-sex test sample (n=50)

	Directional Difference		Absolute Difference			
Variable	Raw diff (kg) Mean (SD)	%PE Mean (SD)	Raw diff (kg) Mean (SD)	|%PE| Mean (SD)	Wilcoxon T-value	20% (%)
BORB	-27.8 (100.2)	-40.0 (154.2)	65.4 (80.3)	92.6 (129.0)	0.06	16
HORB	-839.8 (2246.3)	-1164.2 (3181.2)	898.3 (2223.1)	1241.0 (3151.4)	0.01	4
BIOR	-10.2 (48.4)	-14.5 (71.7)	37.3 (32.1)	53.0 (49.9)	0.14	20
BPOR	-7.2 (50.2)	-6.4 (60.1)	34.1 (37.2)	44.8 (40.1)	0.31	28
LFM	-45.3 (218.9)	-50.3 (241.6)	96.6 (201.3)	117.9 (216.3)	0.15	14
BFM	-120.1 (415.9)	-134.6 (427.4)	172.0 (396.8)	205.3 (397.7)	0.05	12
ORBA1	-39.5 (110.1)	-56.1 (154.7)	75.1 (89.2)	105.8 (125.5)	0.01	20
ORBA2	-41.0 (111.5)	-58.1 (156.7)	76.0 (90.8)	107.1 (127.7)	0.01	20
ORBA3	-38.1 (118.8)	-57.5 (176.2)	77.2 (97.5)	111.4 (147.5)	0.03	19
FMA1	-22.8 (136.3)	-24.2 (145.0)	62.6 (122.9)	76.5 (125.1)	0.24	26
FMA2	-22.9 (136.4)	-24.3 (145.1)	62.6 (123.0)	76.5 (125.2)	0.24	26
FMA3	-11.2 (106.0)	-10.5 (115.8)	55.1 (90.9)	90.9 (93.7)	0.47	24

Table A13. Differences between known and estimated mass for postcranial variables (MA equations), combined-sex test sample (n=50)

	Directional Difference		Absolute Difference			
Postcranial	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)	Wilcoxon T-value	20% (%)
BIB	1.8 (40.5)	2.9 (54.3)	31.2 (25.6)	42.4 (33.5)	0.75	34
FLM	-41.0 (114.3)	-49.7 (137.9)	62.6 (103.9)	77.8 (123.9)	0.01	28
FHB	8.4 (35.3)	13.5 (45.8)	31.1 (18.3)	41.7 (22.5)	0.10	18
FNB	3.3 (37.2)	5.2 (48.6)	27.1 (25.4)	36.9 (31.6)	0.54	28
MLSB80	6.5 (33.2)	11.6 (42.1)	26.3 (20.9)	35.0 (25.7)	0.18	36
MLCB80	-12097.8 (42667.1)	-18919.1 (74958.8)	12156.4 (42650.1)	18997.6 (74938.5)	0.05	0
CA80	0.0 (19.0)	0.6 (24.7)	15.1 (11.4)	20.3 (13.8)	0.99	56
I80	-236.2 (49.8)	-326.2 (74.5)	236.2 (49.8)	326.2 (74.5)	0.00	0

	Directional Difference		Absolute Difference			
Postcranial	**Raw diff (kg) Mean (SD)**	**PE Mean (SD)**	**Raw diff (kg) Mean (SD)**	**|PE| Mean (SD)**	**Wilcoxon T-value**	**20% (%)**
MLSB65	-16.7 (35.9)	-19.2 (43.3)	25.3 (30.3)	32.6 (34.2)	0.00	40
MLCB65	-0.7 (37.1)	-5.6 (50.2)	25.2 (27.0)	33.7 (37.4)	0.89	38
CA65	10.1 (14.5)	13.7 (18.1)	14.9 (9.4)	19.6 (11.2)	0.00	52
I65	-241.0 (48.6)	-332.7 (68.4)	241.0 (48.6)	332.7 (68.4)	0.00	0
MLSB50	6.5 (25.6)	11.7 (31.8)	20.4 (16.6)	27.2 (19.9)	0.08	44
MLCB50	-30.6 (95.1)	-52.5 (165.6)	45.7 (88.7)	69.9 (158.9)	0.03	28
CA50	4.3 (16.9)	6.8 (21.1)	13.3 (11.1)	17.7 (13.2)	0.08	62
I50	-295.7 (71.8)	-404.3 (80.6)	295.7 (71.8)	404.3 (80.6)	0.00	0
MLSB35	7.6 (32.3)	13.6 (39.4)	25.2 (21.3)	33.8 (24.0)	0.10	34
MLCB35	74.6 (18.5)	100 (0.00)	74.6 (18.5)	100 (0.00)	0.00	0
CA35	6.3 (14.9)	9.6 (19.5)	13.7 (8.4)	18.6 (11.1)	0.00	56
I35	-612.1 (143.6)	-837.2 (165.1)	612.1 (143.6)	837.2 (165.1)	0.00	0
MLSB20	-1.7 (67.6)	3.3 (78.1)	46.5 (48.6)	59.4 (50.2)	0.86	20
MLCB20	-6488.2 26922.2	-7475.8 30728.2	6552.4 26906.3	7562.8 30706.5	0.09	6
CA20	9.9 (17.9)	12.8 (23.5)	16.4 (12.0)	22.0 (15.1)	0.00	48
I20	-189.0 (49.0)	-263.2 (76.8)	189.0 (49.0)	263.2 (76.8)	0.00	0

Table A14. Differences between known and estimated mass, cranial variables (RMA equations), combined-sex test sample (n=50)

	Directional Difference		Absolute Difference			
Variable	Raw diff (kg) Mean (SD)	%PE Mean (SD)	Raw diff (kg) Mean (SD)	|%PE| Mean (SD)	Wilcoxon T-value	20% (%)
BORB	1.5 (20.7)	-2.1 (29.0)	15.2 (14.0)	20.8 (20.1)	0.62	52
HORB	1.4 (23.1)	-2.9 (30.4)	17.9 (14.4)	24.5 (17.8)	0.67	46
BIOR	-1.3 (18.5)	-5.6 (26.8)	13.1 (13.0)	18.5 (20.0)	0.62	68
BPOR	2.2 (16.7)	0.4 (21.8)	12.3 (11.4)	16.4 (14.1)	0.35	66
LFM	3.3 (24.4)	0.9 (27.7)	18.1 (16.4)	22.9 (15.2)	0.35	50
BFM	0.5 (20.9)	-3.2 (27.2)	16.4 (12.7)	22.2 (15.9)	0.86	50
ORBA1	1.8 (21.9)	-1.8 (29.8)	16.7 (14.1)	23.0 (18.7)	0.57	56
ORBA2	1.3 (21.9)	-2.5 (30.0)	16.7 (14.1)	23.1 (19.0)	0.69	56
ORBA3	0.3 (23.2)	-3.9 (32.6)	17.7 (14.8)	25.1 (20.8)	0.94	54
FMA1	4.5 (21.5)	2.8 (24.6)	16.6 (14.2)	20.9 (12.9)	0.14	42
FMA2	4.5 (21.5)	2.8 (24.6)	16.6 (14.2)	20.9 (13.0)	0.14	42
FMA3	2.7(22.6)	0.2 (26.5)	17.3 (14.6)	22.3 (13.9)	0.41	40

Table A15. Differences between known and estimated mass, postcranial variables (RMA equations), combined-sex test sample (n=50)

	Directional Difference		Absolute Difference			
Postcranial	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)	Wilcoxon T-value	20% (%)
BIB	9.0 (18.5)	9.3 (24.5)	15.7 (13.2)	20.6 (16.0)	0.00	58
FLM	5.9 (21.1)	5.5 (24.9)	15.5 (15.3)	19.4 (16.3)	0.05	58
FHB	7.4 (17.5)	8.2 (21.2)	15.0 (11.6)	19.2 (11.9)	0.00	58
FNB	-0.2 (19.8)	-2.6 (25.9)	14.9 (12.9)	20.0 (16.5)	0.95	60
MLSB80	3.2 (15.9)	2.9 (20.5)	12.8 (9.8)	17.0 (11.6)	0.16	62
MLCB80	-3.6 (26.9)	-11.2 (40.3)	21.3 (16.5)	30.7 (28.2)	0.34	46
CA80	-8.3 (16.1)	-11.9 (21.7)	13.8 (11.5)	19.2 (15.5)	0.00	62
I80	-469.4 (135.7)	-638.4 (150.6)	469.4 (135.7)	638.4 (150.6)	0.00	0
MLSB65	12.9 (13.1)	16.3 (15.1)	14.7 (11.0)	18.8 (11.7)	0.00	52
MLCB65	-8.4 (29.2)	-16.9 (42.0)	22.3 (20.4)	31.7 (32.1)	0.05	44
CA65	17.0 (12.4)	22.0 (13.4)	17.8 (11.1)	23.1 (11.4)	0.00	44
I65	-504.9 (143.7)	-684.4 (144.5)	504.9 (143.7)	684.4 (144.5)	0.00	0

	Directional Difference		Absolute Difference			
Postcranial	Raw diff (kg) Mean (SD)	PE Mean (SD)	Raw diff (kg) Mean (SD)	|PE| Mean (SD)	Wilcoxon T-value	20% (%)
MLSB50	8.4 (13.3)	10.7 (16.5)	12.8 (9.1)	16.6 (10.4)	0.00	58
MLCB50	-13.7 (32.9)	-25.7 (54.1)	23.7 (26.5)	35.8 (47.8)	0.00	42
CA50	19.6 (11.7)	25.9 (12.6)	19.8 (11.3)	26.2 (12.0)	0.00	28
I50	-995.7 (317.4)	-1338.4 (280.3)	995.7 (317.4)	1338.4 (280.3)	0.00	0
MLSB35	13.2 (13.9)	16.7 (17.5)	15.9 (10.6)	20.8 (12.2)	0.00	50
MLCB35	72.5 (18.5)	97.0 (0.7)	72.5 (18.5)	97.0 (0.7)	0.00	0
CA35	-20.1 (16.3)	-27.3 (21.3)	20.9 (15.3)	28.3 (19.9)	0.00	46
I35	-226.2 (68.1)	-308.0 (78.2)	226.2 (68.1)	308.0 (78.2)	0.00	0
MLSB20	3.0 (20.8)	2.3 (27.5)	17.1 (12.1)	22.8 (15.20	0.31	48
MLCB20	2.3 (22.7)	-1.1 (28.9)	18.0 (13.9)	23.5 (16.5)	0.48	52
CA20	16.4 (14.7)	20.8 (17.6)	18.8 (11.6)	24.3 (12.2)	0.00	38
I20	-268.1 (95.9)	-366.8 (120.7)	268.1 (95.9)	366.8 (120.7)	0.00	0

Table A16. Summary data for the postcranial variables, BMI-Restricted sample

	Female (n=57)		Male (n=59)		Combined-sex (n=116)	
Variable	Mean (SD)	Range	Mean (SD)	Range	Mean (SD)	Range
BIB	275.6 (18.3)	223.1-317.5	280.3 (16.6)	211.2-313.3	278.0 (17.5)	211.2-317.5
FHB	45.4 (2.4)	39.8-55.5	50.5 (2.9)	41.5-56.6	48.0 (3.7)	39.8-56.6

Table A17. Difference between known and estimated mass for post cranial regressions, BMI-restricted samples

	Female (n=57)			Male (n=59)			Combined-sex (n=116)		
Equation	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)	PPE Mean[1] (SD)	|PPE| Mean[2] (SD)	20% (%)
FHB-1	-27.8 (15.6)	-27.9 (15.3)	40.3	-16.3 (11.3)	16.7 (10.5)	61.0	-21.1 (13.6)	21.4 (13.2)	49.1
FHB-2	-5.1 (13.1)	10.6 (9.2)	82.5	-1.9 (9.7)	8.0 (5.8)	98.3	-3.5 (11.6)	9.3 (7.8)	90.5
FHB-3	-13.2 (13.9)	15.2 (11.7)	68.4	-8.7 (10.3)	10.9 (7.9)	84.7	-10.9 (12.4)	13.0 (10.1)	76.7
FHB-4	-7.5 (13.3)	11.4 (10.1)	80.7	-3.9 (10.5)	8.8 (6.8)	91.5	-5.2 (11.8)	9.7 (8.5)	87.9
Mean of FHB 1-3	-15.4 (14.2)	16.1 (12.0)	66.7	-9.0 (10.4)	11.4 (8.1)	81.4	-11.8 (12.5)	13.7 (10.4)	75.9
Mean of FHB 1-4	-13.4 (14.0)	14.6 (11.5)	70.2	-7.7 (10.4)	10.4 (7.7)	84.7	-10.2 (12.3)	12.5 (9.9)	80.2
STBIB-1	-3.5 (10.6)	8.9 (6.6)	93.0	4.3 (8.3)	7.4 (5.7)	96.6	-0.1 (11.0)	8.1 (6.2)	94.0
STBIB-2	-3.2 (10.6)	8.8 (6.5)	87.7	3.1 (8.5)	6.9 (5.7)	96.6	0.5 (10.9)	7.9 (6.2)	93.1

PPE: percent prediction error (known - estimated)/known * 100, |PPE|: absolute percent prediction error, 20%: percent of individuals whose estimated body masses fall within +/-20% of known mass. 1. Directional differences (positive values indicate underestimation, negative values indicate overestimation); 2. Absolute differences; * indicates significance at p=0.01.

Table A18. Correlations (r) between absolute percent prediction error and age for cranial equations (Aiello and Wood LS regressions)

Variable	Female (n=125)	Male (n=128)	Combined-sex (n=253)
BORB	0.159	0.068	0.125
HORB	-0.037	-0.007	-0.037
BIOR	0.262	0.028	0.156
BPOR	0.152	0.049	0.123
LFM	0.074	-0.116	0.002
BFM	0.024	-0.009	0.015
ORBA1	0.039	0.020	0.031

Table A19. Correlations (r) between absolute percent prediction error and age for postcranial equations

Equation	Female (n=125)	Male (n=128)	Combined-sex (n=253)
FHB-1	0.145	-0.060	0.087
FHB-2	0.145	0.005	0.106
FHB-3	0.158	-0.028	0.104
FHB-4	0.063	-0.025	0.105
STBIB-1	0.104	-0.014	0.076
STBIB-2	0.104	-0.019	0.076

Table A20. Comparison of mean percent prediction errors for the cranial variables using the LS regressions derived in the present study, with and without age-restriction

	Non-age restricted (x=46.1 yrs) (n=50)	Age restricted 18-60 (x=43.4 yrs) (n=46)
Cranial Variable	|PPE| Mean[1] (SD)	|PPE| Mean[1] (SD)
BORB	19.2 (15.0)	19.6 (13.8)
HORB	19.8 (15.5)	20.1 (14.4)
BIOR	18.1 (15.4)	18.4 (14.3)
BPOR	17.1 (13.5)	17.7 (13.23)
LFM	18.8 (13.7)	19.1 (12.9)
BFM	19.5 (14.8)	19.9 (14.2)
ORBA1	19.0 (15.2)	19.5 (13.8)
ORBA2	19.0 (15.2)	19.5 (13.8)
ORBA3	19.4 (16.1)	19.6 (14.6)
FMA1	18.8 (14.0)	19.2 (13.4)
FMA2	18.9 (14.1)	19.2 (13.4)
FMA3	18.9 (14.4)	19.3 (13.6)

Table A21. Comparison of mean percent prediction errors for the postcranial variables using the LS regressions derived in the present study, with and without age-restriction

	Non-age restricted (x=46.1 yrs) (n=50)	Age restricted 18-60 (x=43.4 yrs) (n=46)
Postcranial Variable	**|PPE| Mean[1] (SD)**	**|PPE| Mean[1] (SD)**
FHB	18.1 (12.5)	16.1 (10.2)
BIB	16.8 (13.0)	16.7 (11.4)
FNB	16.2 (10.7)	18.2 (11.4)
FLM	18.4 (12.4)	18.7 (11.9)
MLSB80	15.4 (12.0)	15.4 (11.2)
MLCB80	20.6 (18.3)	20.9 (15.0)
CA80	12.8 (10.3)	12.6 (9.5)
I80	144.1 (42.3)	142. 6 (41.4)
MLSB65	13.0 (10.1)	12.6 (8.9)
MLCB65	20.4 (15.5)	20.6 (15.3)
CA65	14.1 (9.9)	13.5 (8.6)
I65	153.6 (41.0)	150.9 (38.8.)
MLSB50	12.3 (10.2)	12.0 (9.4)
MLCB50	20.2 (15.4)	20.7 (15.1)
CA50	11.1 (8.9)	11.0 (8.8)
I50	211.0 (48.4)	208.1 (46.5)
MLSB35	14.7 (10.8)	14.8 (10.9)
MLCB35	20.6 (15.6)	20.9 (14.9)
CA35	11.9 (8.4)	12.3 (8.5)
I35	166.1 (44.6)	165.7 (45.3)
MLSB20	17.2 (12.5)	17.2 (11.8)
MLCB20	21.0 (15.9)	21.2 (15.2)
CA20	16.5 (12.4)	16.5 (12.1)
I20	156.6 (49.4)	156.0 (48.9)