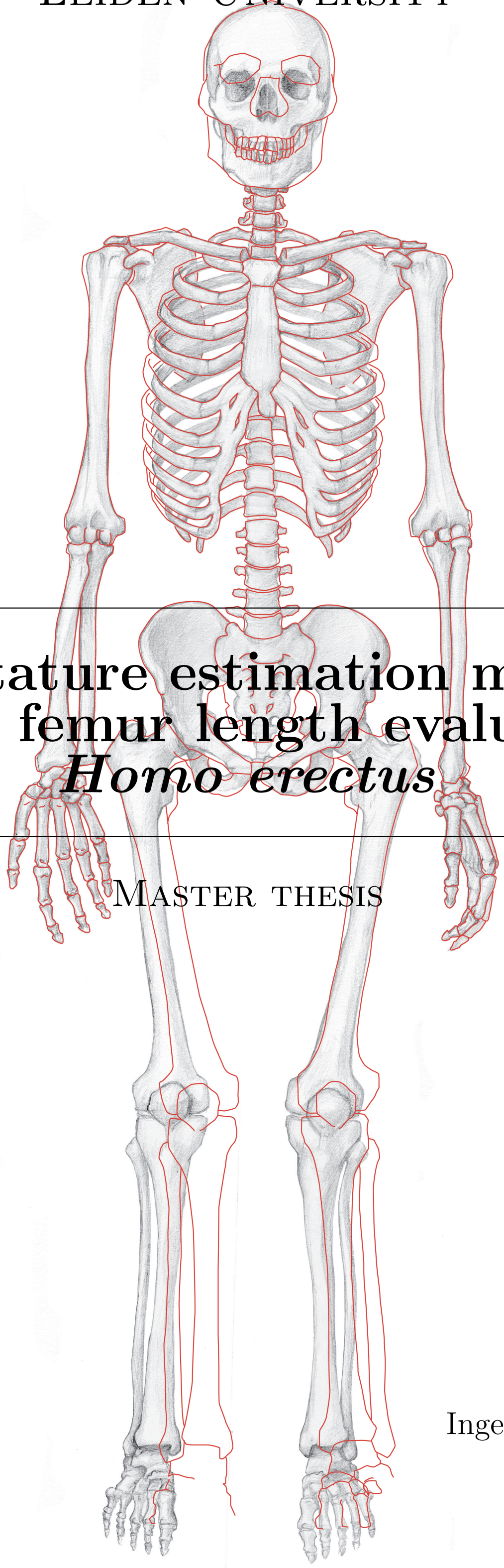
LEIDEN UNIVERSITY



Body stature estimation methods based on femur length evaluated for *Homo erectus*

MASTER THESIS

Inge Silvie van Stokkom

July 27, 2012

Cover figure © 2011, Fabiën Bruine de Bruin. The figure shows a modern human skeleton in black with the skeleton of KNM-WT 15000 as proposed by Ruff and Walker (1993) in red.

Body stature estimation methods based on femur length evaluated for *Homo erectus*

By

Inge Silvie van Stokkom
S0202134

MASTER THESIS

UNIVERSITY OF LEIDEN

FACULTY OF ARCHAEOLOGY

Palaeolithic Archaeology

Supervisors:
Prof. dr. Wil Roebroeks
Dr. Kevin Kuykendall

Leiden, July 27, 2012

Contact:
Inge Silvie van Stokkom
Email: `ISvanStokkom@gmail.com`

Contents

Ik draag deze scriptie op aan mijn oma's,
die beiden op hun manier een voorbeeld voor mij zijn

Preface

In this preface I would like to thank some people who have helped me along the way. As it is a personal message, I address them in the language I use with them.

First of all, I would like to thank my classmates and teachers at the University of Sheffield who all contributed to my fantastic year as an exchange student there. My time here was inspiring and challenging, and I have gained knowledge and understanding from all our discussions and from living as an expat in a foreign country. Thank you for helping me out when I failed to understand the English ways. I hope that some of you will be lasting friends.

It was at Sheffield that I started with this thesis, which is how dr. Kevin Kuykendall became my thesis supervisor, even though I am to be graduated from the University of Leiden. I would like to thank both of my supervisors, next to dr. Kuykendall it was prof. dr. Wil Roebroeks who has guided me, for their flexibility and encouragement. I took the freedom and the time to explore the subject of body size in hominins before I finally settled down to work on the analyses that have been written down here, and they have allowed me this freedom.

Verschillende mensen zijn mijn proeflezer geweest van grote en minder grote stukken van de uiteindelijke tekst. Zij hebben zich veel tijd en moeite getroost om eerst te begrijpen waar ik mee bezig was en me vervolgens van bruikbaar commentaar en discussie te voorzien. Ik wil hiervoor graag de volgende mensen bedanken: Robine Houchin, Willemien Meinders en Eric Wierda. Kathy MacDonald has also been a supplier of very useful comments right at the end when it was almost finished, but there were still some difficult issues to sort out. Thank you!

Twee mensen in het bijzonder hebben me tijdens een groot deel van het proces bijgestaan als ik op momenten niet meer wist hoe het verder moest, omdat ze allebei veel weten van het proces “scriptie schrijven”, hoe dat technisch in z’n werk gaat en wat de regels van de faculteit precies zijn. Deze ondersteuners zijn Kinge Siljee en Femke Tomas. Jullie beiden ook erg bedankt.

Mijn familie, vrienden en huisgenoten hebben de afgelopen jaren vaak met me meegeleefd en me veel bemoedigende woorden gegeven. Door deze steun en warmte is het nooit bij me opgekomen om op te geven. Dank jullie allemaal daarvoor, en speciaal ook aan Fabiën Bruine de Bruin voor het tekenen van de kaftillustratie.

Dan zijn er nog drie mensen zonder wie dit allemaal echt niet gelukt was. De eerste twee zijn mijn ouders, Annemiek en Eelco van Stokkom. Zij hebben niet alleen mijn leven mogelijk gemaakt, maar hebben ook achter elke keuze gestaan en die op alle manieren ondersteund. Ik hoop dat jullie niet teveel slapeloze nachten hebben gehad omdat deze scriptie wat langer duurde dan gedacht. Bedankt allebei!

De laatste die ik graag wil bedanken is Roel Jordans. Naast mijzelf is er niemand die zo betrokken is geweest bij dit project, als technische ondersteuner (ik was zonder hem nooit uit LaTeX gekomen), proeflezer, wiskundedocent en sparringpartner. Daarnaast had ik zonder hem een heleboel meer stress gehad, want ik ken niemand die altijd zo rustig blijft en dit gevoel ook zo weet over te dragen als hij. Dank je wel voor al je hulp op alle fronten en ik denk dat onze weekenden vanaf nu een stuk flexibeler in te delen zullen zijn.

Bedankt allemaal!

Inge van Stokkom

Chapter 1

Introduction

1.1 Body size of *Homo erectus*

1.1.1 Introduction

The study presented in this MA thesis is concerned with the estimation of the body size during lifetime of hominin individuals commonly assigned to the species *Homo erectus*. Body size has a meaningful role in the biology of a species and is correlated with many variables of behaviour, life history et cetera (Foley 1987; Schmidt-Nielsen 1984). In extinct species, these kinds of variables cannot be observed directly. However, their relationship to physical variables like body size makes it possible to reconstruct them. Body size in humans refers to the length of the body in standing position from souls to the top of the head, or stature, and to a human's weight or mass. In this thesis, the body size discussed will be stature.

Known physical attributes of *Homo erectus* have been used to predict or infer unknown behavioural and physiological attributes. Examples of this are home range size, necessary daily caloric intake, and social system, but also environment and food type (Aiello and Key 2002; Antón 2003; Foley 1987). Variables that can be used as predictors are not only body size, as mentioned above, but also body proportions and known environmental and behavioural attributes such as the vegetation type at *Homo erectus* sites and the tools found there.

Estimations published in the 1990's (McHenry 1991a, 1992; Ruff and Walker 1993) form the basis of often-quoted average body sizes for hominin species (Aiello and Key 2002). A picture has emerged of a clear distinction between the australopithecine group and *Homo erectus* in overall species mean stature, body proportions and body mass sexual dimorphism. The observed trends have therefore been used to construct life history characteristics for both groups (McHenry and Coffing 2000). The consensus view brought forward here was that australopithecines had a distinct, and now extinct, life pattern of their own, while *Homo erectus* was the first hominin to show a human-like life pattern and body proportions together with a distinctly bigger brain, indeed making him the first true *'Homo'* (McHenry and Coffing 2000; Wood and Collard 1999). This image of *Homo erectus* has been influenced heavily by the find of a unique and very complete skeleton, KNM-ER 15000 (also known as Nariokotome boy or Turkana boy) (Ruff and Walker 1993). It was predicted that KNM-ER 15000, a juvenile, would have measured 1.85 m and weighed 68 kg as an adult (Ruff and Walker 1993), which would have made him taller than members of previous hominin species hitherto found and among the taller individuals of *Homo sapiens* alive today.

New fossils have been found since the publications of McHenry (McHenry 1992) and Ruff and Walker (Ruff and Walker 1993). These fossils show sometimes surprising morphology, provenance, dating, or all of these combined. Some of these specimens have been estimated to be on the short and light side of the *Homo erectus* range, like the Dmanisi specimens (Lordkipanidze *et al.* 2007). Since fossils that can be used for body size estimation are rare, the addition of several of these new finds to the existing pool could influence the average and range for *Homo erectus* body size greatly. This raises the question of the inferences on behaviour and ecology that were based on the older finds are still valid or whether the image we have of *Homo erectus* biology needs to be adjusted.

To evaluate whether newly found fossils change the inferences that can be made for *Homo erectus*, it is necessary to compile a database with all available fossils for body size estimation and use these fossils to calculate the range and average of the species and search for patterns (e.g. geographical) within these data. However, the practice of estimating body size is not without controversy and there are different methods available. To properly weigh the different individual estimates, the estimation methodology of body size itself has to be examined. In this thesis an attempt will be made to evaluate the existing methods in use for the estimation of body size, in order to assess the reliability and accuracy of their findings. This will result in an overview of the possibilities and constraints of the estimation of body size.

1.1.2 Definition of the problem

The answer to the question of early hominin body size depends on the method used to calculate it. The importance of the question and the uncertainty of the level of correctness of the answers, have urged researchers to find different methods for estimating that size. In the search for the best predictor equations, different reference samples, statistical methods and other potentially influential variables have been used. The results of these different methods vary, even for the same fossil. When establishing the average size of *Homo erectus*, a choice needs to be made between the available estimations and estimation methods. To aid in this choice, this thesis will try to establish an answer to the question of which body size estimation technique is most reliable for *Homo erectus* by evaluating these estimation techniques. Once it has been established which methods are the most reliable, it will also be clear which results are best used in the calculation of species averages and of species biology.

The relevance of this research lies in the evaluation of the factors that influence the results of body size estimation. As averages of body size estimations are being used to calculate other aspects of a species' biology, these estimates partly determine our conception of the lifestyle of *Homo erectus*. Body size estimates should therefore be as accurate as possible. The reliability of the estimation methods will be assessed by evaluating the methods on the quality of their variables (reference sample etc.) in Chapter 4. The suitability with *Homo erectus* will be assessed by an analysis of the factors that influence the result of an estimation method in Chapter 5.

This thesis will present an overview of the different estimation methods and their variables. The results of these methods are presented per fossil in the appendix. These data will help to establish whether some methods are more reliable than others and, consequently, which estimation results are best used. This evaluation of the estimation techniques will focus on the reliability and accuracy of the factors that form the estimation equation: the reference sample used, the variables measured and the mathematical technique used for calculation. This is further explained in Chapter 2 and operationalized in Chapter 3.

Due to reasons of scope, the focus will lie on the evaluation of methods that estimate body stature and will not evaluate body mass. Although body mass is the body size variable that is most often used in biology to study other biological aspects of a species, stature is in bipedal animal also a very relevant characteristic. An example of its usefulness is the assessment of the amount of sexual size dimorphism within a species with it (Aiello and Key 2002; Carlson *et al.* 2007).

Within the available body stature estimation methods, the focus has been placed on methods that use femur length only as a predictors. This focus has been chosen as femur length is one of the variables that is most commonly used in body stature prediction and one of the most highly correlated variables with body stature Hauser *et al.* (2005); Lundy (1985); Trotter and Gleser (1952). A number of fossil lengths have been collected from the literature. These are all femora. Although the focus is on *Homo erectus* fossils from a wider taxonomic background have been included in the study. The large size range of these fossils allows for a broader study of the characteristics of the estimation techniques. Four of the fossils have been chosen to do statistical tests on, these are OH 28, KNM-ER 1463, D4167 and Berg Aukas. These represent the smallest (KNM-ER 1463), the largest (Berg Aukas), a modern human-sized one (OH 28) and an intermediate one, on the edge of the human femur range, (D4167) of the sample here collected (see the appendix).

1.1.3 This thesis

This thesis consists of 8 chapters. The first two present an introduction to the subject (Chapter 1) and general considerations concerning the estimation of stature (Chapter 2). Chapter 3 presents the materials and methods used in the analyses. Chapters 4, 5, and 6 all present the results of one approach to the research question. Chapter 4 contains a discussion on the quality of the reference samples that are used for stature estimation. Chapter 5 analyses the different variables ("factors") that influence the stature estimates from the equations. Chapter 6 presents an approach to estimate the stature of three hominin case-studies, AL 288-1, KNM-WT 15000 and D4167. For overview purposes, a subject discussion is given in these respective result-presenting chapters. A more general discussion is presented in Chapter 7, where the results will be placed in a larger framework and connected to the larger questions on *Homo erectus* body size. A conclusion will be reached in Chapter 8.

1.2 Body sizes

There are different kinds of size measurements that can be used when evaluating the body size of *Homo erectus*. They are the ones which we use in describing ourselves too, as they are length and weight also known as stature and mass. Another, less commonly used, measurement is width, usually taken to be bi-iliac breadth. Stature and mass of a *Homo erectus* individual are the two objectives of body size estimation. They are correlated to one another ($r = 0.7$ in modern humans, see Ruff and Walker 1993) and are both considered to be useful characteristics to aid in the understanding of *Homo erectus* biology. Body width is not an estimation objective itself, but is used as a predictor variable together with stature to estimate mass (Ruff *et al.* 1994). This section will discuss the different measurements of body size and the reason for its importance.

1.2.1 The importance of body size

This subsection will give a short introduction to the functioning of body size in relation to other lifestyle variables and also to the way in which these dependent variables are calculated by using known or calculated body mass. The other chapters will subsequently focus on the methods of estimation of body size.

Ruff (2002, 211–212) made a list of reasons why it is important to appreciate body size and body size variation in extinct species.

- Body size is related to life history parameters, ecology, social organisation and is often used to predict these traits in fossil taxa.

- Body size is the usual denominator for assessing key evolutionary trends i.e. encephalization, megadontia, robusticity, gut size.
- Body shape is a taxonomic and population marker.
- Body size and shape of earlier humans can serve as effective baselines for assessing more recent temporal and geographical variation.

As Ruff states in this list, body size is not only important for making inferences on the life history and other variables of an extinct species, but also for evaluating evolutionary trends. Brain size, for example, is seen as an important characteristic of any extinct hominin. Brain size is usually discussed as relative brain size to body size (EQ) rather than absolute brain size. Incorrect estimations of body size can thus create a false image of the mental capabilities of a species, since relative brain size increase is often taken as a proxy of increasing intelligence. A certain range of variation in body size and shape is seen as an evolutionarily meaningful, adaptive, characteristic of a species which is thus informative of taxonomic boundaries. Shifting body size and shape through time is therefore indicative of shifting adaptive pressures.

The estimation of body size in extinct species is important because it is such a central aspect in an animal's adaptation, as it seems related to most ecological, physiological, life history and behavioural parameters ("lifestyle") in many mammals (Feldesman *et al.* 1990; Hens *et al.* 2000; Schmidt-Nielsen 1984). The way in which these lifestyle variables influenced extinct hominins is of interest to the palaeoanthropologist. As body size is a central variable in the web of interdependent biological variables of a species, the study of body size is of interest. Indeed, body size is so central to the functioning of an individual (species) that it is impossible to treat all of its consequences within the space of this thesis

Body size determines the broad ecological niche position that a species can fill. For example, almost all nocturnal primates weigh less than 1 kg, and insectivorous primates weigh less than 250 grams, whereas folivorous primates are usually much larger (Aiello 1992b). Diet and period of activity, related to body size, limit the niche possibilities in the ecological community of the species. The example given is a result of the relationship between energy requirements and body weight, which in mammals is described by Kleiber's law $BMR = kWb^{0,75}$. [BMR = basal metabolic rate, kWb = energy consumption]. From this law follows that larger mammals consume more absolute energy, while smaller mammals consume more energy per unit body weight (hence eating more energy-rich food, like insects) (Aiello 1992b). This basic relationship allows by inference for the identification of broad lifestyle patterns in extinct primates.

Life history variables are closely related to body size, especially so in subfamilies of primates (Charnov 1993; Charnov and Berrigan 1993; Harvey and Clutton-Brock 1985; Robson and Wood 2008). This relationship can be quantified in extant species and used to predict these life history (and physiological, behavioural, and ecological) variables from body size in extinct species. Body size is correlated with the mentioned aspects of adaptations within species to such a degree that Damuth and MacFadden (1990) consider body size to be the most useful single predictor of the adaptations of a mammal ("single" referring to the possibility that combined traits might be better predictors).

In line with this kind of research there are many publications that try to quantify hominin lifestyle and relate it to body size. A prime example of this is Aiello and Key (2002), who interpret the large size difference between, especially, females of early *Homo* and *Homo erectus* in terms of the energetics requirements that a larger body size carries with it, and with this the increasing energetics costs of having offspring. They postulate that this energetic burden is so high, that social changes could have been necessary to cope with it Aiello and Key (2002).

These biological characters are however mostly researched in relation to body mass, not body stature. In biology, it is generally mass that is measured and used and not stature, as it is proven that mass is correlated with many life history and other traits (Damuth and MacFadden 1990; Foley 1987). Stature is usually, though not always, ignored because most mammals are not bipedal and thus their body height (which is not really "stature") is of little biological relevance. Stature in the knuckle-walkers is measurable and these measurements are sometimes reported. In bipedal animals, like all species in the genus *Homo*, stature is a relevant biological attribute which could be informative on the lifestyle of the species and gives a suggestion of the size of an animal that we can intuitively get a grasp on (Hens *et al.* 2000; Nakatsukasa *et al.* 2007). Stature influences body mass in bipedal animals Hens *et al.* (2000). Mass and stature are correlated to each other so a mass estimate can be calculated from a stature estimate to give an indication of mass, this procedure will gain very large error ranges as the correlation between both body sizes is only .7 (Ruff and Walker 1993). However, as Feldesman *et al.* (1990) states, *"body weight is commonly regarded as the best measure of body size"*.

Both stature and mass are thus characteristics that can be useful for estimating the body size of hominins, and it will depend on circumstances, such as the condition of the fossil itself or the research goal, if both or just one of them can be estimated. For example, when taking a perspective on extinct hominins from the primate order, mass is the more useful estimation goal as this is what is measured in the primate reference sample and can thus be compared. Stature estimation from a bone has been conceived as useful for a long time already, for example for forensic or purely scientific goals.

The earliest work on establishing the relationship between body size of an modern human individual and its bones started already in the 18th century, though came under more systematic study in the late 19th century. Rollet (1888) collected a sample of 100 French cadavers on which he published measurements, subsequently used to explore stature prediction for a given bone length by Manouvrier (1892). The use of statistics in this field was introduced by Pearson in 1899 (Pearson 1899), who used linear regression to obtain formulae which could be used for estimation. Pearson was also the first to explicitly state that data from one 'local race' may not be suitable to fit another of such a local race, as *"stature is quite as marked a racial character as cephalic index"*. He mentions the Palaeolithic in this context: *"the extension of the stature regression formulae from one local race - say, modern French - to other races - say Palaeolithic man - must be made with very great caution"* (Pearson 1899, 177). As an aside, the word "races" used here should, more than a century later, rather be read as "populations", or "peoples", Pearson (1899) in the example cited above mentions "the French race".

Subsequently, other investigators turned to other populations to make stature equations (Stevenson 1929; Telkkä 1950) and undertook large-scale data collection (Breitinger 1937). Regression was the method used in all these investigations. In modern times, data collection in diverse populations continued (Allbrook 1961; Feldesman and Lundy 1988; Maijanen and Niskanen 2010; Radoinova *et al.* 2002; Ross and Manneschi 2011; Trotter and Gleser 1952), often to be used in forensic cases (Hauser *et al.* 2005; Wilson *et al.* 2010). The methods have not changed much over the years, measurements on bones are still taken with the use of an osteological board (Hauser *et al.* 2005, 187). Although new measurement techniques using radiography or MRI are explored as well (Pelin *et al.* 2010).

Mass and stature together provide a more complete picture of the body size on an individual, or a species. An example for this is given by Aiello (1992a) for AL 288-1 or "Lucy". Aiello shows that the mass and stature of this individual were unknown in this combination for any human, i.e. no human of her stature had the mass that was predicted for Lucy. Such a picture illustrates in what ways extinct hominins can be different from the variation modern humans show today.

1.2.2 Body stature

Body length or stature is the length of the total body in standing position, measured from the soles of the feet to the uppermost part of the skull. The bones that contribute to it are the calcaneus, talus, tibia, femur, pelvis, vertebrae and cranium (Raxter *et al.* 2006). A smaller contribution comes from the inter-vertebral disks, skin and muscles. Maximum stature is obtained in an individual's life after cessation of long bone growth, in modern humans in their early 20s. Stature in this thesis will be given in centimetres.

In modern humans, stature attained in life is influenced by both genetic and environmental factors. It is known that stature is in part heritable. The heritability has been established around .8 (Hernández *et al.* 1998, 549), or for European men between .87–.93, and for European women .68–.84 / .89–.93 (Silventoinen *et al.* 2003). The environmental aspect of size determines how much of the genetic potential for total stature is attained. Stature attained is negatively affected by lack of nutrition, lack of sleep and disease (Hernández *et al.* 1998, 549; Subramanian *et al.* 2010). Stature is therefore proposed to be determined by a complicated interplay between both genetic factors and ontogenetic circumstances that influence the stature attained by any individual.

Stature is by some considered to be more easily and reliably estimated than body mass, because it is more straightforwardly related to the fossil bones found (Feldesman and Lundy 1988). The length of a femur, for example, contributed to the stature of the individual during life, while its weight, which in life contributed to body mass, can have changed due to post-depositional processes. Furthermore, a femur's weight is a smaller part of the body's weight than its length is of body length. A femur, while weighing just a few kilograms, is around 27% of total body stature in humans, 25.5% in bonobo's, 24% in chimpanzees and 23.5% in gorilla's (human ratio see Feldesman *et al.* 1990; non-human ratios see Hens *et al.* 2000). On the other hand, the unknown body proportions of extinct hominins make stature estimation more difficult than it is in modern humans.

There are three pathways that can be followed to estimate stature in life from a fossil. The first method is Fully's anatomical method (Fully 1956; Fully and Pineau 1960), which can be applied to (partially) complete skeletons. In the case of a complete skeleton, body length can be calculated by adding the lengths of the lower limb to that of the axial skeleton bones (Lundy 1985). A correction factor for the spinal curvatures and non-bone body parts can be applied to raise skeletal length to body length. There are actual several of such anatomical methods, but they give very similar results (Maijanen 2009).

Fully's soft tissue correction index is still commonly used for this purpose, although warnings about its dependability apply (Porter 2002), not in the last place because the method is based on a small, modern reference sample. In any case, because of the rarity of nearly complete skeletons, the method is seldom applied in palaeoanthropology. A notable exception is AL 288-1, to which fossil it was performed successfully by Schmid (1983), in Geissmann (1986a).

Stature can also be estimated by using the percentage that a femur takes up of total stature in modern individuals. This is referred to as the "ratio method" (Porter 2002), as it is the ratio of femur to stature that is used. This method has been applied since the 19th century (Rollet 1888). Feldesman and colleagues (Feldesman 1992; Feldesman *et al.* 1989, 1990) have brought it forward again in recent years, with data suggesting that the femur/stature ratio is relatively stable interpopulationally in modern humans (Feldesman *et al.* 1990) and can successfully be applied to extinct hominins.

A third possibility is a regression of stature on a skeletal variable or the other way around (Konigsberg *et al.* 1998). This method is the most commonly used. Limb bone length and body length are highly correlated in modern humans and especially the bones of the lower limbs are suitable for length estimation (Hens *et al.* 2000; Trotter and Gleser 1952, 1958). This method has been applied since 1899 when Pearson introduced it (Pearson 1899). The result of the regression analysis is a formula, which can be used to calculate the body size from a skeletal variable.

Several methods are thus available for stature estimation from bones. While the anatomical method is the most reliable because there is no dependence on a correlation in a modern reference sample, it can hardly ever be applied. For fossils that preserve too little material for the anatomical method to be applied, the other two are necessary. From a mathematical perspective the ratio method is a special kind of regression equation; thus, these two methods are not strictly separated (Hens *et al.* 2000). As Pearson (1899) explains, a linear regression equation is more sensitive to the population it is based on and estimates the statures at the small and large side of the population more accurately than the ratio from the same population. This is due to the usage of an intercept in the regression formulae, while it is emitted in a ratio.

1.2.3 Body mass

The body mass of an individual is its weight in kilograms. As we all know from personal experience, body mass can and will change during life whereas stature is more fixed after reaching maturity. Mass responds to food availability and quality, disease and pregnancy. One individual can have had a broad range of body masses over its lifetime. This intra-individual variability should result in larger error margins in estimations for body mass than in stature estimation. The prediction will for a large part be dependent on the average "body mass state" of the reference sample it is based on; obese, athletic, or underweight. The preferred choice for reference samples is usually athletic or "average" individuals, and excludes heavily obese or emaciated individuals.

There are several types of methods used for estimating body mass; the main divide is between the so-called mechanical and morphometric approaches (Auerbach and Ruff 2004). Mechanical methods use the correlation between certain skeletal measurements and body size to estimate body size from these measurements. Morphometric methods use the shape relationships of the body (e.g. between stature and mass) to estimate mass. There is no body mass equivalent to Fully's anatomical method for stature reconstruction.

In a species performing bipedal locomotion, pressure generated by the body mass or body weight is directed through the lower limbs (Aiello and Dean 1990). The diameter of the long bones of these limbs and the surface size of the joints need to be able to carry the full body weight and are thus expected to vary with this weight. Therefore, it is the proximal and distal femoral articular areas, proximal tibial joint, mid-diaphysis diameter of long bones and last lumbar vertebra that are most often used to estimate body weight (Damuth and MacFadden 1990; Ruff *et al.* 1997). Methods that employ these estimators are grouped as mechanical methods (Auerbach and Ruff 2004).

Morphometric methods try to establish body shape and size directly from the bones. Mass, for example, can be estimated from stature alone but also, and more accurately, from stature and body width (bi-iliac pelvic breadth). In this way, the shape of the body is taken into account (Auerbach and Ruff 2004). Measurements on modern populations are used to create multiple regressions of body mass on stature and bi-iliac breadth. Mass can thus be predicted from the combination of length and pelvic breadth (Ruff *et al.* 1997; Ruff and Walker 1993). Stature and mass have a correlation coefficient of 0.7 in modern populations, which is generally regarded as not high enough to allow reliable estimations to be made from stature only (Ruff and Walker 1993). By incorporating another linear measurement of body volume, 2 out of 3 volumetric measurements are used, which increases confidence in the results.

The body mass index method (BMI) (Porter 1999, 2002) can be viewed as a morphometric method, as it uses likely shape at given stature to calculate body mass. The variables used in BMI-calculations are stature, mass, age and sex. When stature, sex and developmental stage can be determined on a skeleton, likely mass can be inferred. Its use in palaeoanthropology is unfortunately limited due to lack of relatively complete fossil skeletons, with known stature, sex and age, and uncertainty about the BMI-patterns in extinct species.

The analysis of hominin body mass cannot be executed in the same style as the analysis of hominin stature. Many formulae are available for the estimation of stature from femur length as there is a forensic interest in the reconstruction of stature from body parts. Stature is one of the characteristics which helps identifying a corpse with unknown identity, while mass is not used in this way (probably because it is estimated too inaccurately). Since stature is estimated most accurately when the estimation formula is based on the population the individual belonged to, equations from many populations are now available. Body mass has not seen such an interest due to limited relevancy in present crime-solving, and equations estimating mass from skeletal elements are consequently less abundant. They stem from the fields of biology and palaeoanthropology itself, e.g. Hartwig-Scherer (1993); McHenry (1991b, 1992); Ruff (1990); Ruff *et al.* (1997); Steudel (1980).

Due to the low number of available body mass estimation equations, it is not possible to analyse them in a similar fashion to the stature estimation equations also due to lack of time. However useful body mass estimation may be in *Homo erectus* research, it will be not regarded in this thesis and only the estimation of stature will be investigated. As the two kinds of body size are closely related, advancement of knowledge on one part can help to elucidate the other as well. Even though the correlation might be too low to calculate one from the other, it is clear that a taller individual will be heavier, even if it is not proportionate.

Chapter 2

Background to body size estimation

In this chapter the factors that influence body size estimation will be discussed. A body size estimation method is grounded in the data on which it is calculated. The different variables and data that are needed ("factors") to calculate an estimation equation will be assessed in this chapter, in section 2.1. The mathematics employed to execute the calculation of the actual estimation equation from the data are discussed in section 2.2. They are discussed in order to be able to better understand and evaluate the results.

The goal of this stature estimation evaluation is to be able to estimate the stature of *Homo erectus* more reliably and accurately. When it is understood how the factors of stature estimation function in interplay, a more informed choice of a technique to be used on *Homo erectus* can be made. A short introduction on *Homo erectus* is therefore also provided in section 2.3. It will discuss the *Homo erectus* femoral raw material that is available to do stature estimation on.

Information on the functioning of these factors in estimation will help in the evaluation of the different techniques that employ these factors in drawing up body size estimates. This evaluation will be done using three approaches in Chapters 4, 5 and 6. The approaches will be explained in Chapter 3, Material and Methods.

2.1 Variable considerations

A body size estimation equation is constructed by a regression technique from stature and predictor variable size data of a given reference sample. The ingredients needed for the construction of a regression equation for estimating body size are therefore a predictor variable, a reference sample measurement set and a mathematical technique. Here, the input variables that are needed for the creation of a formula are discussed, namely, predictor variable and reference sample. The predictor variable should be a variable present and measurable in fossils. It should furthermore be measured in a reference sample that consists of individuals of known body stature and mass. The reference sample data is subsequently used to calculate the correlation coefficient and the formula, which can be used to estimate the body size for a certain predictor measurement.

As there are many estimation equations (e.g., 110 have been collected in the dataset used in this thesis), the reference sample and predictor data have been varied and extracted from different sources. Not all these sources are of the same quality, and a low quality dataset will likely affect the quality of the regression equation created from that dataset. Furthermore, it is likely that some of these reference samples will be better fitted for the estimation of extinct hominins than others. In the following sections, the requirements that can be set for reference sample and predictor data will be discussed. This will help identifying data which are better than other data, and which will be more likely to give reliable estimation equations. The considerations on what would be a suitable reference sample for *Homo erectus* will also be presented.

2.1.1 Reference samples

Correlation between variables can only be determined in one or more extant species, which will form the reference sample. This correlation can never be known for extinct species, which is why it is important that the reference sample resembles the extinct species or individual as closely as possible in correlation coefficients between predictors and body size and in body size itself. The more the size of the fossil deviates from the mean of the reference sample, the more inaccurate and unreliable the estimations will be (Damuth and MacFadden 1990; Kurki *et al.* 2010). However, it is difficult to assess how much deviation, and in what characteristics (size, shape), is present between the reference sample and the extinct species (Konigsberg *et al.* 1998). In addition to these problems, the reference sample data itself may have issues that affect their reliability and usability.

Requirements for reference samples

There are several issues with reference samples which can influence the results of the methods that are based on them. The data are subject to certain requirements, intended to guarantee their reliability for the purpose of body size estimation, that are hardly ever met completely (Lundy 1985; Porter 2002). Data needed for body size estimation is generally not easily available and for some of the large skeleton collections the dependability of the records is questionable (Steudel 1980).

Reference sample data should ideally conform to a set of requirements, the most detailed set of requirements listed by Porter (2002, 23–24). The requirements that are listed here are derived from Porter with some adjustments made. They are less elaborate as the requirements that cannot be checked with the dataset at hand are excluded from the list. Porter's design is to outline the requirements of any new study, the objective of this thesis is to test the ones that exist for reliability.

These requirements are on several levels. On the level of the population, care should be taken to restrict influences from mingling or migration which could possibly mix people with different genetic or phenotypic geographical adaptation. The subjects should come from a homogenous population with a clear and known history, preferably with low incidence of intermixture with other groups that are genetically separated in deep time. Homogeneity of a population can be seen from different angles (e.g. genetic, historic, cultural, environmental) and is always relative because variation between people is present in the smallest group. The homogeneity that should be optimised is dependent on the exact study objective, and the homogeneity present in a population will determine which level of comparison can be achieved. However, it has been known for a long time already (Pearson 1899) that an estimation equation from one population can give inaccurate results for another population, so it seems prudent to try to separate such populations. Because populations share so much variables (most importantly, genes and environment), homogeneity can be seen as the similarity of these factors and measured on a very crude level by known migrations.

On the level of the individual, an effort needs to be made to ensure that measurements are as accurate as possible, restricting other possible influences by measuring in a uniform way, e.g. by removal of footwear, using the same equipment and procedures. Stature and mass of the live subjects should be recorded according to a strictly prescribed set of measurements and, to avoid inter-observer variability, by the same observer. The subjects should have a normal health and should not be obese or emaciated, and their age should be known and should follow the distribution of the age-groups in the population (e.g. the sample should not be made up exclusively of people aged over fifty). The subjects should preferably have been measured within a few years before their death. After death, the identity of the subjects should never be lost. Data on stature and body mass during life should be associated with the specific predictor measurements of the individual.

Although it would be useful to study obese and emaciated individuals to gain insight in the variation range, especially for forensic cases, they are better excluded from reference samples to be used for palaeoanthropological goals because they suffer from an unhealthy condition that probably, especially for the obese situation, had a much lower incidence in earlier hominin populations than in populations today due to differences in living conditions. Subjects with extreme body weight will not only influence the equations on body mass, but also the equations used for stature estimation. Moreover, there are indications that emaciated subjects have lost body length after attaining adult stature under normal health (Porter 2002, 19).

The third level of requirements is that of the predictor measurements. These should also be measured according to a strictly prescribed set of measurements, in dry state, again by the same observer. It is preferable to measure the predictor that will be used, in the case of stature estimation the long bone predictor, instead of calculating it from another measurement like limb segments.

A list of specific indications of lowered reliability of a reference sample can be drawn from these considerations.

- A low correlation coefficient, for example by the use of predictors that have a low correlation with body size; e.g. upper limb instead of lower limb bones (Trotter and Gleser 1958), or living limb segment measurements instead of bone measurements.
- Small reference sample sizes, a minimum of 30 to 50 individuals is taken as a standard. Pearson prefers even more: *"When the correlations of the organs under consideration are high (e.g. long bones in Man), fifty to a hundred individuals may be sufficient; in other cases it is desirable that several hundred at least should be measured"*. (Pearson 1899, 170).
- Unreliable data for the reference sample, e.g. unclear measurement sets, body size measurements recorded long before death or taken from statements of relatives (Wilson *et al.* 2010)
- Unassociated body size and bone measurements (not from the same person).
- Unrepresentative samples, e.g. only old, diseased or individuals of a certain social class are included.
- Unclear population history and/or large-scale mixing of populations makes it more difficult to match such a population to a hominin individual in terms of, e.g., geography.
- Averaging equations from different populations and sexes.
- Extrapolation necessary due to differences in size between the reference sample and goal individual.
- Compound error by estimating femur length from a femur fragment and from that body size, both by regression.

The ideal requirements are hardly ever met in real life. Many modern groups have not retained genetic isolation from other groups, with large-scale intercontinental mixing starting already in 1500 AD. The data of the reference sample should preferably consist of associated measurements of living mass, living stature and the relevant skeletal predictor elements made on the same individual. This is not always possible. For most reference samples it is the case that either the skeletal elements and mass and stature are measured at death, or they are all measured during life. While measuring during life increases the measurement error of the skeletal element, measuring after death increases the chance of unrealistic mass and stature measurements (Steudel 1980). Hardly any reliable records of live body size can be combined with actual bone measurements (Wilson *et al.* 2010).

From these considerations follow, that many requirements exist that would make a reference sample more reliable for size estimation. Summarized and divided into three categories they are presented in the list below. They will subsequently be discussed in the following subsections.

- Measurements: bone or limb segment; stature (in vivo/on corpse/Fully's)
- Reference sample: sample size; reliable data; associated data; age distribution
- Fossil match: average body size; average femur size

Measurements

To minimize noise, it is recommended that the condition of the reference population is as similar to the that of the goal individual as possible. This not only goes for general body size and body proportions, but also for e.g., the condition in which the bones are measured (preferably dry), age and sex of the individuals, measurement of body size during life, etcetera. The leading principle is to come as close to the goal individual as possible as this will minimize the possible errors (Hens *et al.* 2000).

A measurement to be evaluated is the exact predictor variable that is measured; is this e.g. the femur itself (preferably in dry state) or are living limb segments measured from which the bone length is calculated? The first option is preferable since it is exactly what can be measured on the fossil bone, if complete. The limb segments obviously correlate strongly with femur length (Olivier 1976) but error is increased because bone cannot be measured without increasing the measurement by measuring the skin and other soft tissue. So in limb length data collected on live subjects it is the true measurement of the femur that needs to be estimated. When a regression formulae is used to calculate femur length from limb segment length, and from that stature, compound error increases the confidence interval.

The other measurement that needs to be ascertained is the stature that is measured. The goal is to estimate body size during life, so it is preferable when the reference sample subjects are measured during life. Often however, body size is measured upon death on the corpse. Death changes stature as it relaxes muscles and so stature at death is supposed to be greater than during life (Hauser *et al.* 2005; Pearson 1899). Corpse measurements are again preferable to measurements on skeletons, as skeletons are even further away from the living body that is to be estimated. Allowances for soft tissue need to be made, usually using Fully's method (Fully 1956; Fully and Pineau 1960), which increases the possible error (Porter 2002).

The following table 2.1 presents a list of measurements which can be measured on the reference sample. If there is a preference for a certain recording option when the reference sample is used for size estimation on a fossil, this is indicated with an "X". There is no preference indicated for all of measurements, as in some cases e.g. bone side it is simply better to match than to have a preference; some fossil bones are right and some are from the left side.

Table 2.1: Measurement choices for reference sample variables

Measurement level	**Options**	**Preference**
Bone length	actual bone limb segments in vivo	X
Femur length	maximum length physiological length	
Bone state	fresh with cartilage fresh without cartilage dry with cartilage dry without cartilage	 X
Bone side	right left	
Stature	on living individual on corpse in supine position on corpse in suspended position on skeleton according to Fully's method	X
Individual	age sex	
Individual measurement association	bone length - body size	

However, many of the reference samples that are used are already measured and published. To make them suitable to match otherwise measured individuals, correction factors have been calculated. Some of these are presented in the table below. There is quite some disagreement in the size of the allowances that need to be made, e.g. between Manouvrier (1892) and Pearson (1899) on fresh vs. dry bone. Not a lot of research has been done to accurately calculate these allowances. Pearson (1899) only measured three bones which he manipulated into several states of dryness. Fully too had a small reference sample which was badly controlled (Porter 2002). The suitability and accuracy of these allowances are thus questionable.

Next to this, it seems (e.g. Pearson1899) that the difference it makes in the final result are not very large especially when compared to the error ranges that accompany any estimation using a modern reference sample for a fossil individual. Therefore, and for reasons of scope, no allowances are used in the formula calculations in Chapter 5, even when this is specifically recommended in the publication of the formula. When allowances would have been calculated, this no doubt would have given more accurate representations of the formulae which recommend them, but it was unfortunately not possible.

No allowances will be calculated but also no effort will be made to otherwise match the state of measurement in the reference sample to that of the hominin reality when comparing the results of the estimation techniques (Chapter 5). This means that only one femur length entry per fossil as published will be used, both in formulae which use physiological and maximum length, in formulae which are calculated on right or left sided bones, and in formulae of which the reference sample stature has been measured either on living individuals, corpses or skeletons. However, some of these measurement states will be used for reference sample suitability evaluation in Chapter 4. Only the states of which one of the options is preferable over the others are used for this. For example, reference sample stature measurement state can be used for evaluation but bone side is not.

Table 2.2: Recommendations for the condition of the measurements

Measurement compensation	**Recommendation**	**Author**
From fresh bone to fossil state (dry bone)	subtract 2 mm for cartilage allowance	Manouvrier
	subtract 7.1 mm for cartilage and animal matter allowance	Pearson (1899)
From left to right side		Pearson (1899)
From corpse stature to living stature	add 1.26 mm (M) or 2 cm (F)	Pearson (1899)
From skeleton to living stature	add 10 cm	
To compensate for old age shrinkage	add 3 cm	Pearson (1899)

Reference sample

Some reference samples have been collected some decades ago already, and only the skeletons remain. Complete skeletons can be reconstructed according to Fully's anatomical method (Fully 1956) or the adjusted Fully's method (Raxter *et al.* 2006). Calculated skeleton height needs to be corrected to give living stature. A correction factor of 10 cm is applied to skeleton under 153.5 cm, 10.5 cm for skeletons between 153.6–165.4 cm, and 11.5 cm for heights exceeding 165.4 cm (Feldesman and Lundy 1988; Fully and Pineau 1960). These correction factors proposed by Fully have been criticised due to the reference sample it was based on; World War II concentration camp victims who were probably emaciated (Porter 2002, 19). They are however still applied widely (e.g. Feldesman and Lundy 1988; Formicola 1993).

Criteria to evaluate the reference sample used are the sample size, as small sample size increases the possibility of chance dictating the formula and thus missing the general trend in the population. Data also need to be reliable and preferably associated, meaning that the data on body size and predictor variable come from the same individual. The age distribution of the sample is a minor criterion, important only is avoiding primarily senescent or adolescent samples.

For many of the reference samples that would be interesting to use in palaeoanthropological research (e.g. small-bodied humans and nonhuman hominoids), no or few data are available (Kurki *et al.* 2010). Within the human sample, it is clear that a lot more data are available for certain populations or regions (e.g. Europeans) whereas for other populations no or few data are available (see Chapters 5, 7). Some of the reference samples are based on only few individuals. Although a number of 30 to 50 is usually regarded as the required minimum to rule out the effect of chance, groups with less than 10 individuals are also used (e.g. Pygmy samples). This is an effect of low availability of measurements, for example due to low research intensity or small population sizes. The nonhuman hominoid groups suffer from the same problem. It is considered best to include only wild-shot individuals, to avoid any influence due to the conditions of captivity, and these are quite rare to come by.

Body mass is another matter entirely, as the predictor measurements used for body mass estimation, joint size and cross-sections, can hardly be measured reliably on live subjects. Records from the time of death need to be used, as these are usually the only available indications of subject's body mass. They cannot be checked via a method like anatomical reconstruction because such a method does not exist for body mass. Subjects may often have experienced a change in body mass (weight loss) prior to death, for example when they died because of an illness. Body mass recorded at death may therefore not be representative for body mass during life (Steudel 1980).

Selection of a reference sample for a hominin stature estimation

Just as the condition in which the reference sample is measured is recommended to be as close to the condition of the goal (hominin) individual as possible to minimize noise, it will be beneficial for the estimation if the reference sample will be as similar to the goal individual in other aspects as well. In this section, it will be discussed how this similarity between reference sample and hominin can be approached.

The usage of reference samples in hominin size estimation is hardly avoidable, since almost all methods depend on it. When an estimation equation is applied to a hominin individual, the level of accuracy of the result is dependent on the similarity in size, shape and correlations between the fossil and the reference sample (Kurki *et al.* 2010). As these characteristics are unknown before estimation for almost every hominin fossil, it is difficult to ascertain which criteria should be used for selecting the best reference sample. Certain fossil elements can be used as an indication of full body size. Evidence from the length of the fossil femur shows that the length of some hominin individuals must have been well below the average of modern humans; e.g. the fossil individual with a femur of 28 cm will have a shorter stature than the modern human average with a femur of around 45 cm. When equations based on modern humans are applied on such fossils, extrapolation outside the size range of the reference sample takes place. Since it cannot be predicted with any degree of certainty how the relationship between predictor and size develops where there is no sample data, this greatly increases the error range (Aiello 1992a; Hens *et al.* 2000).

Modern reference samples are never identical to the extinct species and many of the different size-shape forms possible may be seriously under-represented today due to differential extinction (Damuth and MacFadden 1990; Plavcan *et al.* 2002). Having perfect models for all modern species, if this were even possible, is of no help when it comes to examining an extinct species with entirely different body proportions, niche and environment, especially when the species is specialised (e.g. *Paranthropus*). No matter which species are used as reference species, none of these is a perfect fit with any extinct species. Depending on which hominin species is being studied, either nonhuman ape or human body proportions and size might be more similar to those of the extinct species. When selecting the material for the reference sample it is attempted to approach the extinct species as closely as possible, in phylogeny, adaptations (morphological-functional group), and allometric proportions (Damuth and MacFadden 1990). In doing so, *"the most relevant biological context"* should be our leading principle according to (Hens *et al.* 2000, 769). These aspects will be discussed below.

Phylogeny is stressed because it increases the chances of shared inherited stable traits and correlations. For hominins, the extant hominoids are most often opted for, but sometimes a broader sample of primate species is used. As Steudel (1980, 63) writes, *"a variable that has a very high correlation with size across this range of primates is likely to be effective also in* *the prediction of size among fossil hominoids".* The exclusive usage of modern humans is also reported, especially for hominins that belong to the genus *Homo.* The human species has so much internal variability that sometimes only certain populations are used as a reference sample, like small-bodied populations (Feldesman and Lundy 1988; McHenry 1992).

Shared adaptations between reference sample and goal species increase the chances of similar functional relationships. For example, the ecogeographical adaptation hypothesis that Ruff (Ruff 2002; Ruff and Walker 1993) has put forward posits that climate and environment shape human body proportions, both in modern humans and in extinct species. According to him, the climate-related adaptations that modern humans show are also visible within *Homo erectus* and therefore it is best to use modern reference samples coming from the same latitude (as a proxy for climate).

Likewise, overall body size that is similar between reference species and goal species increases the chances of shared allometry and reduces the need for extrapolation outside the size range of the sample. Extrapolation should be avoided because there are no indications how the relationship between the predictor variable and body size will behave outside the size range of the reference sample (Aiello 1992a). The similarity in size should be assessed by reporting the "observed range" of size in the reference sample.

The match between reference sample and extinct species is seen as less of a problem in *Homo erectus* than in earlier hominins such as the australopithecines. There is only one phylogenetically appropriate bipedal reference species, which is the extant *Homo sapiens.* Modern humans share the mode of locomotion (obligatory bipedalism) with *Homo erectus* (McHenry and Coffing 2000). As for example Hens *et al.* (2000) show for KNM-WT 15000, it is likely that *Homo erectus* was quite similar in size and shape to modern humans (see also Ruff and Walker 1993). This however, does not eliminate the theoretical problem of the unknowability of any extinct species so care must always be taken not to equate an entire extinct species with an extant species, especially on the basis of a few only partially complete fossil skeletons separated in time and space. It should be added that modern humans are less robust than these early representatives of our genus (Ruff and Walker 1993, 255). Geissmann (1986a) considers this as evidence that actually *Homo sapiens* is not a mechanically appropriate reference sample for earlier hominins, even when mode of locomotion is shared.

Body proportions different from *Homo sapiens* are expected in the fossil record of pre-*Homo* hominins, but should also be expected in all *Homo* species. Intraspecific differences in body proportions can be observed in modern human populations today, as proportions adapted to climate and environment (Ruff 1994). We are aware of the ecogeographical patterning visible when we compare Inuit and Masai, and likewise recognise the different body composition of our cousin (sub)species, the Neanderthal. It is important that the researcher takes the possibility of such intraspecific differences into account when predicting body size regardless of the taxonomic status of the specimen under consideration.

As intraspecific variability in general size and in proportions in modern humans is high, the results of regression equations calculated on one modern human population will be different from the results of regression equations calculated on another modern human population (Pearson 1899; Stevenson 1929; Trotter and Gleser 1952, 1958). Within a population, stature and body proportions can even change with new generations (Trotter and Gleser 1951). An informed choice thus still needs to be made from the available reference samples, even when making estimations for a modern human (Trotter 1970). From this it follows that formulae based on modern humans are not directly applicable to other members of our genus either. However, as all modern human populations are (probably) similarly related to Early Pleistocene *Homo erectus*, there are no obvious reference populations. The choice made is often one of convenience, by choosing a reference sample that is easily available (e.g., the equations from Trotter 1970; Trotter and Gleser 1952, 1958). Ruff and Walker (1993), already mentioned above, chose a population with body proportions that were assumed to be similar to those of KNM-WT 15000 because of hypothesised shared eco-geographical adaptations (hyper-tropical body proportions). Others (Olivier *et al.* 1978) made the choice on body size, by taking Pygmies as a reference sample for expected small-bodied australopithecines.

However iconic, KNM-WT 15000 is just one individual found at a certain geographical area. The exact body proportions of all individuals assigned to *Homo erectus sensu lato* are not known and since it is such a broad species in time, geography and morphology, there is no guarantee that conclusions can be extended from one fossil to another or to the whole species across its entire period. Using a broader reference sample of e.g. hominoid species can still be helpful, as their body proportions differ from those of modern humans. This is especially useful in australopithecines and early (transitional) *Homo*, who have body proportions unlike those of modern humans (McHenry and Coffing 2000). Even if the results obtained from using these reference samples give very unrealistic results for a particular type of fossil, at the very least their irrelevance will have been established. Differences in the body size estimates based on different parts of the body can also be informative on the body proportions of the fossil (Fortelius 1990), as for example McHenry (1992) shows in his careful assessment of body mass predictions for several bone elements of the same individual. A relatively long humerus to body stature will result in very large estimates when an equation based on standard modern human body proportions is used, thereby informing on the size of the humerus relative to stature.

2.1.2 Predictor variables

The reliability of any estimation of an animal's body size depends upon the selection of those variables that are most strongly correlated with body size (Smith 2002). There is some discussion on what constitutes a good variable. The general criterion is the amount of correlation between the variable and the measure of body size that is estimated. Correlation coefficients that are considered reliable for the purpose of predicting should have a value above $r > 0.8$. However, correlation coefficients above 0.9 or 0.95 exist, and these are preferred when available (Steudel 1980). Grine *et al.* (1995) on the other hand, accepts low correlations ($r = 0.57$) for multiple correlations in intraspecific samples. Another criterion is the availability of the variable both in the fossil material and in measurements of the reference sample. A very high correlation coefficient is of no use when the predictor bone is never found in the fossil record.

Fossil material can be divided into two groups that need different treatment; cranial and postcranial material. Postcranial material is often considered as more reliable for size estimation than cranial material, because there is a functional relationship between the postcranial skeleton and size (Ruff *et al.* 1997). Even so, several researchers have presented methods that use cranial measurements to estimate body size of fossil individuals (Aiello and Wood 1994; Kappelman 1996; Spocter and Manger 2007; Steudel 1980). It is the correlation coefficient with body size that determines the usability of a certain predictor, and not its functional relation with size (Smith 2002, 274). Of course, the existence of a functional relation with size would help to attain a high correlation coefficient, but even when such a relationship is absent the variable is not useless for prediction. Cranial material has an advantage over postcranial material in that the taxonomical definitions are commonly based on cranial material and thus morphology. Crania are therefore more easily assigned to a taxon, while isolated postcrania are difficult to assign (Antón 2003).

Obviously, as we are reconstructing an extinct species, the exact correlation of any variable with body size in that particular species is not known. The fact that the correlation coefficients must be calculated using one or several modern species potentially introduces error, as the correlation present in modern species might differ from that the extinct one. A functional relationship between the variable and body size in this context gives extra credibility to the assumption that indeed the correlation coefficient was high in the extinct species too.

A special cranial variable group is the dental one. Molar occusal surfaces are often used for body size estimation in other primates (Fortelius 1990). This is not considered a good method in the hominin lineage because there is strong dental surface increase in the *Paranthropus* line, while in the *Homo* line there is dental reduction (McHenry and Coffing 2000). Both changes are thought to be related to a diet switch; a clear example of a changed function of the variable. This weakens the inherited pattern present in less-specialised primates between the variable (molar in this case) and body size. Other cranial measurements, however, may retain high correlations with body size. Using non-functionally related predictors in extinct species, however, does increase the likelihood of error as there the species might have had a deviant correlation compared to the general phylogenetic trend.

The problem of unknown correlations in extinct species, however, might be pressing in hominin cranial material. Some of the cranial variables show great diversity in hominins and even within the *H. erectus s.l.* group. It is likely that some cranial variables have been under evolutionary selection in the hominin lineage, as there is a pattern of increasing cranial capacity over time visible in the *Homo erectus* group (Antón 2008). Moreover, there is evidence that some cranial variables are not independent of cranial size (Antón 2008). This means that body size estimations based on these variables will fluctuate with cranial size. In the same research, Anton et al. showed that body size did not show a significant increase pattern over time within *Homo erectus*. Cranial capacity is thus changing independently of body size and variables not independent of cranial capacity might be bad predictors of body size.

Not all postcrania have equally high correlation coefficients and some are more responsive to other factors during life besides size. Research on sheep with different activity levels (Lieberman *et al.* 2001) showed that articular surfaces give a more reliable estimate for body mass than diaphyseal diameter, as the last variable is more responsive to mechanical loading (e.g. stress on the bones caused by activity). The articular surfaces are influenced to a smaller (non-significant) degree by loading and are therefore more purely reflective of the permanent loading factor that is body mass (Lieberman *et al.* 2001).

Both cranial and postcranial material are used for body size estimations. The femur (femur length) in particular has been used for regression equations to estimate stature, though also the tibia, the spine or a combination of these can be used (Feldesman and Lundy 1988). The femur has also been much used in body mass estimation, especially the femur head. This thesis will focus on stature estimation methods based on the femur. This choice has been made because most size estimation techniques are based on the femur, as this has proven to be the most reliable stature predictor in modern humans (Hauser *et al.* 2005, 189; Lundy 1985, 74).

2.2 Mathematics

2.2.1 Regression

In addition to the problem of the choice of data to be used in calculations, there is also that of the mathematics underlying these calculations. The method most often used in body size estimation is linear regression, both for stature and for mass. Different paths can be followed in these regression calculations, with each method requiring different elements to be calculated from the dataset. The estimations that result from these different approaches can be different too; thus, the choice of which technique is used has influence on the final result. How do these differences arise and how big is this influence? Which regression technique would be best suitable and give the most reliable results?

A regression formula is a description of a relation between two or more variables. If body size (y) is partially explained by the size of a bone (x), this relationship can be mathematically described if it is known to what extent the size of x predicts y. If the body size is a stature, this relationship can be a linear function, because body size in this case is the same dimension when the variable is bone length size. Its form is the commonly known $y = ax + b$. This will become a power function when variables of different dimensionality, like mass (volume) and the length or surface of a bone, are compared. Mass estimations are therefore treated on a logarithmic scale, usually by log-transforming the data of both axes. Both natural logarithms and base-10 logarithms can be applied. Biological, allometric, relationships are usually linear when log-transformed, and linear regression can be applied to the data in this form.

In order to create an estimation equation, first the reference sample dataset needs to be analysed using regression analysis. The data of the reference sample are used as input to calculate the formula which describes the relationship between the variable and size in that sample. The data are visualised in a bivariate graph. Methods of regression are then used to find the appropriate r and function to describe the relationship. The three main techniques used for this are ordinary least square (OLS), major axis (MA) or reduced major axis (RMA) analysis (Aiello 1992a; McHenry 1992). The differences between these line-fitting techniques lie in their usage of the characteristics of the data (correlation coefficient, ratio of the variances).

OLS is a Model I approach; the other two are Model II approaches. The least squares techniques that are used are inverse calibration and classical calibration (Hens *et al.* 2000). Model I methods are asymmetrical, meaning that regressing y on x (e.g. stature on long bone size) gives another result than x on y (e.g. long bone size on stature). This results from the division of variables used into either the dependent (y) or independent (x; known) variable of which only the variance of the last one is used in the calculation: OLS analysis uses the covariance of x and y divided by the variance of x. The difference between inverse calibration and classical calibration is this assignment of (in)dependence to the different variables by the researcher. Classical calibration extends the useful range of variation around the mean stature and compensates for the differences in scaling proportions (Hens *et al.* 2000). Data on which OLS is employed should answer to certain assumptions. The independent variable should be sampled without error and the reference sample data should be normally distributed.

Model II regression methods differ from Model I regression methods because they do not divide the variables into dependent and independent variables. This means that Model II methods are symmetrical. There is no assumption that either x influences y or the other way around, while it is assumed that both variables are measured with error (Hens *et al.*

2000). Major axis (MA) and reduced major axis (RMA) are both Model II approaches. RMA is a compromise between the inverse calibration and classical calibration lines, RMA extends the useful range of variation around the mean stature compared to the inverse line, but not to the same extent as the classical line (Hens *et al.* 2000). MA assumes equal error variances for the long bones and size (Hens *et al.* 2000).

The statistical method belonging to the femur/stature ratio method (Feldesman *et al.* 1990) can also be named the zero intercept regression ratio model (Hens *et al.* 2000). Its mathematical description is a simple $y = ax$, lacking the intercept b. After all, the length of a femur is multiplied by the stature/femur ratio to obtain the stature estimate and no further adjustments are made. Hens *et al.* (2000, 781) call this *"a special case of classical calibration"*. The ratio model assumes isometry - similar body proportions for all possible statures. When the variables are not scaled isometrically with stature the estimate will be biased (Hens *et al.* 2000).

McHenry (1991a) points out that the correlations obtained with the OSL, MA and RMA methods are so strong that all of these methods can be said to lead to the same conclusions; hence, that it does not matter which method is used. Hens *et al.* (2000) also conclude that the differences between the stature estimates obtained by the different methods are small. However, others (Aiello 1992a) prefer RMA, because this method is never affected by changes in the scale of the variables or the correlation coefficient of the reference sample. Aiello also observes that while OLS might be the best method when prediction is the objective, RMA is to be preferred in cases where the strict conditions for OLS are not met. Moreover, she considers RMA to be more suitable in situations where extrapolation is necessary. Maijanen and Niskanen (2010), too, conclude RMA to be more suitable than OLS for both the tall and short people in a population. Hens *et al.* (2000) on the other hand, see classical calibration (OSL) as the most reliable choice, even in such cases (likewise does Smith (2009)). Even so, several authors give the results for all three or two of the regression techniques (Feldesman and Lundy 1988; Hens *et al.* 2000; Jungers 1988; Maijanen and Niskanen 2010). Most other authors of stature estimation equation publications however, only use OLS and do not further elaborate on their choice for this technique.

Extrapolation and allometry

The allometric formula is $Y = kX^a$ (x is body size, y is the size of the character, k is allometric coefficient and a is the allometric exponent), and when converted to a log-linear relationship $\log Y = a \log X + \log k$ (Aiello 1992b). Positive allometry reflects the increasingly greater proportion that a certain variable is of body size with increasing body size ($slope(a) \geq 1$), negative allometry consequently reflects the decreasing proportion that the variable is of body size with increasing body size (though it still increases in an absolute sense; $0 < a < 1$). Inverse allometry ($a < 0$) and independence ($b = 0$) also occur. In arithmetic scale, allometric relationships are graphically plotted as a curved, non-linear, line. The curve (slope) reflects the changing relationship that the variable has to size. A linear prediction formula can therefore never be identical to an allometric formula.

Extrapolation, that is projecting the regression line out of the size range of the reference sample, should be avoided where possible. It should be borne in mind that whenever regression analysis is used on a reference sample, the resulting correlation line is valid only within the size range of the sample because outside of the range the relationship between the variables may follow a different pattern; this is especially true when the relationship is allometric. The further the regression line is from the mean of the sample, the wider the confidence intervals get (Feldesman and Lundy 1988). However, when there are no phylogenetically and functio-morphologically close, extant populations in the size range of a certain extinct species, it is necessary to extrapolate, as is done, for example, when estimating the body size of australopithecines (Aiello 1992a).

Transformation of data

Data can be used in logged (natural log or log10) or 'raw' (not logged) form. Log is used to transform the data in order to improve the normality of the sample and standardize variance (correct for heterogeneity in variance). Heterogeneity of variance is caused by large differences between the data, for instance when they consist of the body size measurements of all mammals as a class (Kappelman 1996). Log is used in mass estimation, but hardly ever in stature estimation. Where some see the unlogged data as the "real data" (Packard *et al.* 2010), others insist that the logged data are just as "real" and much more useful in research, especially allometric research (Kerkhoff and Enquist 2009). This is because the log-scale makes it possible to compare the relative variation within species of very different sizes instead of the absolute variation (which will be, for a similar relative variation, much larger in an absolutely larger animal than in a smaller animal species). Log transformation is thus used to *"uniformly depict relative variation (..), normalize sample variance and reduce the influence of outliers"* (Kerkhoff and Enquist 2009, 519).

The logged data is subjected to linear regression analysis. This requires several assumptions, e.g. that the logged y is normally distributed. When body size needs to be predicted, absolute data are required instead of just relative variance. This may result in the logged data being detransformed after calculation. These transformations can bias data, this is called the detransformation bias. This necessitates a correction factor to be applied to the (un-logged) results (Smith 1993).

Log-transformation is usually not applied in stature estimation equations. In the dataset collected here (see appendix), only one of the publications (Jungers 1988) makes use of log transformation. It is a very common transformation in body mass research.

2.2.2 Accuracy

Any prediction from a single or from multiple fossil skeletal variables that is based on data from extant species is not going to be exact (Smith 1996), even if sometimes the estimation results are published as a single number with decimals. There is always a spread around the estimated number. After all, there is also spread of measured data points around the regression line in the reference sample ("residuals"). Prediction errors (PE and % PE) and standard errors of the estimate (SEE and % SEE) give an indication of the uncertainty of the model. The general rule is to give a 95% confidence interval (CI), which is twice the two-tailed standard deviation. This gives a certainty of 95% that the estimate falls within the ranges of the CI.

Tests of accuracy used in modern human stature estimation cannot be used for fossil hominins, as they require a sample with known body size. These error measures and the correlation between traits, although known for the population on which the equations were built, cannot be assumed to hold in another population (Smith 1996, 454). However, when the correlation of the reference sample is used for predicting stature in extinct hominins, measurements of error are included as well and perhaps they can be used as an indication of the minimum error to be expected.

The accuracy of an individual hominin is compromised further than the accuracy of an individual belong to the reference population in several ways. First of all is the already mentioned unknown amount of difference between the species of the equation and the species of the estimation. Secondly, there is the often unknown sex of the fossil individual. Modern humans are, of course, easily divided into females and males, and as there are clear size differences between these two groups within a given population, it is obvious that two regression equations need to be made. Fossils can often not be assigned to either sex with any confidence, and the averaging of the female and male results that is done increases the error and invalidates the confidence limits (Feldesman *et al.* 1990, 584).

Thirdly, is there the aspect of compound error as fossil femora tend to be distorted, damaged or otherwise incomplete. Only a few of the fossils are actually complete, and most need some reconstruction. Some of the femurs used in stature reconstruction are only represented by small bone fragments; of others, only the femoral head is preserved. In the case where these parts are used as input in regression equations whereby the total femur length is calculated, a confidence interval needs to be attached to the result. The error that is associated with calculating total length from small parts of the femur can be large (e.g. can exceed 10% of predicted bone length). Additional error is independently added with the next calculation step, which is that of the stature estimation. This compound error makes for very large confidence intervals, and thus for less reliable and potentially even useless stature estimations (Feldesman *et al.* 1989, 584).

The confidence interval of the modern human population can due to these extra sources of error only be seen as an absolute minimum (Trotter and Gleser 1958, 115). The error is much likely to be larger, perhaps even to such an extend as to give error ranges that are so wide that the estimates become meaningless (Feldesman *et al.* 1990; Smith 1996). A last issue to recognize here is that an individual needs to be estimated with a wider confidence interval than an average of a population or a species (Smith 1996) for example due to the possibilities of the influence of personal circumstances during ontogeny, which are averaged out in populations.

The range of error in fossil estimation cannot be measured the way it is measured in modern human research (e.g. Maijanen and Niskanen (2010); Raxter *et al.* (2008)), so it can only be attempted to be minimized by choosing the reference sample that is most similar to *Homo erectus* in all aspects of body size, body proportions, biology and adaptations thus minimizing potential sources of error. These problems might enforce a reconsideration of the aspect of accuracy when stature estimation is applied to fossil hominins. What would be an acceptable deviation in the estimation of the actual stature of an individual? Accepted standard errors in modern human samples are, for example, 4 cm for the Caucasian males in (Trotter and Gleser 1958, 84) and around 3 cm for Polish males in (Hauser *et al.* 2005, 188). What would be considered as an accurate estimate in hominins, if it still needs to carry biological significance? This question needs to be answered by the researchers who use body size estimates in subsequent calculations of e.g. energetic expenditure. In the few test-cases of hominin stature estimation that are given in this thesis (Chapter 6), any estimate within a 10 cm range around the “true stature” of the individual (5 cm both ways) is considered an accurate result, and a 20 cm range as a fairly good result. This however does not say anything about the confidence intervals that need to be taken into account with fossils whose “true stature” is not known, or about the usefulness of such estimates in further research.

Large confidence limits are a necessity when estimating a species with another species estimation equations, but when they get too large they will be meaningless. The sources of possible error associated with this endeavour would caution to take wider confidence limits in mind than when applied to a recent human. It argues against the usage of the exact estimate that is the result of an equation. The use of a hominin estimate might be limited to rough estimates of size, not so much to estimate an individual precisely, but more to identify the larger patterns of stature change through evolutionary time.

2.3 *Homo erectus*

Stature estimation of hominins are used to help to understand the evolutionary patterns in the hominin lineage (McHenry and Coffing 2000). It is the objective of this thesis to evaluate the stature estimation methods available for *Homo erectus*, in order to be able to make more precise and accurate stature estimations in further research. A new *Homo erectus* stature average or new stature estimate for all *Homo erectus* femora will not be presented here as it is not in the scope of the present research. Therefore, a detailed discussion on the Pleistocene postcranial fossils will unfortunately not be given, but a short review on *Homo erectus* in the light of stature estimation will be.

The origin and development of our genus are still so much covered in mysteries, that every new find might stir up long-held assumptions, e.g. (Ferring *et al.* 2011; Potts *et al.* 2004; Simpson *et al.* 2008; Spoor *et al.* 2007). The development of early *Homo* and *Homo erectus* is far from resolved (McHenry and Coffing 2000), especially now *Homo habilis* might be definitely be crossed out as anagenetic ancestor for *Homo erectus* (Spoor *et al.* 2007) and *Homo* individuals with small bodies (Gona at ± 0.9–1.4 mya) and small crania (KNM-OL 45500 at ± 0.95 mya, KNM-ER 42700 at 1.55 mya, the Dmanisi crania at 1.78 mya) remain to exist even long after the first large femora have appeared in the fossil record (e.g. KNM-ER 736 at ± 1.7 mya (Geissmann 1986b) and possible *Homo erectus* KNM-ER 3228 at 1.95 my (Antón 2003)) (Antón 2003; Potts *et al.* 2004; Rightmire *et al.* 2006; Simpson *et al.* 2008; Spoor *et al.* 2007).

In this thesis, it is the *sensu latu* view that is taken of *Homo erectus*, following Gilbert (2008). In this view, *Homo erectus* develops in Africa around 1.8 mya and is represented by there by the early fossils of KNM-ER 3733 and 3883. The species spreads out to Dmanisi (Georgia, 1.78 mya, Ferring *et al.* (2011); Lordkipanidze *et al.* (2007); Rightmire *et al.* (2006)) and Indonesia (1.8 mya for Mojokerto, 1.66 mya for Sangiran; though these dates have been contested, Antón (2003); Swisher III *et al.* (1994)), and is present in China (Zhoukoudian, Hexian) in the Middle Pleistocene (e.g. Dubois 1894; Weidenreich 1939 in Conroy 2005). The species is still present in Africa around 1 mya, as testified by the Daka and Buia crania, after which time period derived forms of humans develop (Gilbert 2008, 360). The late Middle Pleistocene and Late Pleistocene Indonesian finds (Ngandong, Sambungmacan) might or might not (Antón 2003; Gilbert 2008) belong to *Homo erectus*.

These first and last appearance dates (FAD and LAD) give an indication of the temporal range of a species, but likely underestimate this range (Robson and Wood 2008). Most individuals of a species were not preserved to become found fossils after their death. How easily estimations for a temporal range can change is shown by the dating of Out of Africa I (by *Homo erectus*), which was changed from approximately 1 mya to 1.8 mya as a result of re-dating of an existing Indonesian fossil and of the find of 1.78 my old fossils in Dmanisi, Georgia (Lordkipanidze *et al.* 2007).

The large geographical spread combined with the osteological differences between the *Homo erectus* fossils has led some researchers to assign them to different species (Tattersall 2007; Tattersall and Mowbray 2005; Wood 2010). While these divisions might, or might not, reflect the biological situation of that time better than one taxonomic unit that encompasses a large geographical, temporal and osteological spread, the broader usage of the term *Homo erectus* is preferred here. For one, this grouping into a single taxon might actually better reflect the real biological situation after all (Antón *et al.* 2007; Asfaw and et al 2002; Gilbert 2008; Kramer 1993). It is perhaps unsurprising for such a long-ranged and geographically widespread species that there is considerable morphological diversity, including stronger morphological similarity between specimens from the same area and period (Gilbert 2008).

Moreover, establishing actual biological species (according to the biological species concept) in the fossil record is difficult and entails philosophical problems, the discussion of which goes beyond the aim of this study, which is not in the first place taxonomic. Therefore the term *Homo erectus* is used in a loose sense to include fossils that at one point in time have been dubbed "*Homo erectus*" (including *Homo ergaster*) and might or might not have been in the same biological species as the other fossils that are denoted by the term.

Nonetheless, it is recognized that there are morphological differences between geographical groups of *Homo erectus* specimens, which others might call species, and that these groups might be seen as allotaxa (Antón 2003, 154) (i.e., not as reproductionally separated, but with distinctive morphological features) even when they are not recognized as such in the text. However, some of the characteristics that are used to define these groups have been shown to be related to brain size, which increases over time in the taxon (Antón *et al.* 2007). This makes them less usable as a criterion for taxonomic classification, since they are not apomorphies but more strongly expressed general characteristics due to larger brain size. These debates strengthens the choice to consider a very broad view of the species in this study and review the fossils as individuals instead of grouping the results into (sub)species, OTUs or palaeodemes.

This is not to say that taxonomy is an unimportant issue. To be able to synthesize data from individual fossils, it is necessary to group them taxonomically. The choice for a reference sample should be based on the similarities of the reference sample with the extinct species. Our knowledge of (the morphology of) an extinct species is based on the pieced-together information from the fossils found, and the taxonomic groups created are based on the patterns visible in the fossils. These specific patterns are taken into consideration when making a choice for a certain reference sample. For body size estimation, it is therefore best to have taxonomic groupings of fossil individuals that share the same body proportions and absolute body size spread and variation, and, of course, to use the taxonomic grouping that is closest to the real biological situation at one point in time. The taxonomic grouping used will partly determine the patterns that are identified, as parts of the total variation are transferred to another taxonomic group when a fossil is included in that other group (Antón 2003, 152), *"the allocation of individual fossils to each hominin taxon determines the inferences drawn about the life history of that taxon"* (Robson and Wood 2008, 406).

The species is defined mostly by its cranial characteristics. The holotype is the Trinil 2 calotte, found in the 19th century on Java by Eugène Dubois. *Homo erectus* is characterized by general robustness and hypertrophy of (cranial) bone, and by further morphologic characters listed in, e.g. Antón (2003).

2.3.1 *Homo erectus* femora

Hominin fossils are seldom complete, undistorted and unweathered; the measurements made on these fossils from the input of size estimation formulae. If a fossil is distorted, there is a substantial possibility of a wrong measurement being used as input, resulting in a misleading size estimation. Especially for cranial material the accurate measurement is paramount, as here the distances between the measuring points are sometimes only a few millimetres. Thus, a slight miss-measurement may lead to large differences in the final result (see for example Spocter and Manger 2007). The correct reconstruction of a distorted or incomplete fossil is therefore a necessary first step in a reliable body size estimation.

As the taxon *Homo erectus* like other hominin taxa, is mainly defined on cranial material, problems arise when it is the characteristics of the postcranial skeleton of the species that is sought after. Most fossils are found unassociated with cranial remains. It can be difficult to assign such unassociated postcranial bones to a species, especially if there are no known partial skeletons for one of the species linking the cranial morphology to the postcranial morphology. Another level of difficulty is added when there were more than one hominin species living in that area around the time to which a certain postcranial element is dated (McHenry and Coffing 2000). Similarities in postcranial morphology between species complicates the matter further.

The postcranial skeleton of *Homo erectus* is known directly only from the partial skeletons KNM-ER 803, KNM-ER 1808 and KNM-WT 15000 (Antón 2003). From these three fossils it has been concluded that the postcranial bones of *Homo erectus* are characterized by thick cortical bone and general robustness. Of earlier species (sometimes) included in the genus *Homo*, like *H. habilis*, the postcrania are only known from very fragmented remains, from which not many morphological conclusions can be drawn. Derived postcranial characteristics in *Homo erectus* might therefore already be present in earlier *Homo* (if indeed *Homo*, Wood *et al.* (1998)), and are seen as derived from the australopithicine state (Antón 2003). Their characteristics are (Antón 2003, 147), known especially from KNM-WT 15000: *"enlarged articular surface areas of long bones, thick cortical bone (particularly in the lower limb shafts), deep trochlea of the distal femur, double meniscal attachments of the proximal tibia, narrow pelves with marked iliac pillars (i.e., acetabulocrisal buttresses), and medial torsion of the ischial tuberosity"*.

They are retained in later *Homo* species, but some have been lost in *Homo sapiens*

(Antón 2003). This retainment makes it difficult to distinguish *Homo erectus* remains from those of later *Homo* specimens. It has been proposed that the derived *Homo erectus* femur is characterized by its being relatively flattened anterioposteriorly (platymeria), while the tibia is relatively flattened mediolaterally (platycnemia) (Wood and Collard 1999). Conroy (2005, 394) mentions the following defining characteristics of *Homo erectus* lower limb morphology: *"a large acetabulum, a robust vertical iliac pillar, a small auricular surface, a femur characterized by large femoral head, platymeric shaft, thickened cortical bone shaft, narrow meduallry cavity, and heavy muscular markings"*.

In broader perspective, it has been suggested that there is a clear size, and behavioural, distinction between the australopithecine group (including *H. habilis*) on one hand and *Homo erectus* and later species of *Homo* on the other (McHenry and Coffing 2000). The australopithecines are assigned a body size of, e.g. 30 kg and 115 cm (F) and 41 kg and 138 cm (M) for *A. africanus*, whereas for *Homo erectus* (in McHenry and Coffing (2000) *"Homo ergaster"*) averages have been proposed of 56kg and 160 cm (F) and 66 kg and 180 cm (M). It is on the ground of these estimated sizes that unassociated postcranials are sometimes attributed to *Homo erectus* and that postcrania belonging to "large-bodied hominins", like the earliest possible *Homo erectus* postcranials, are thought to be *Homo erectus* (Antón 2003; Antón *et al.* 2007). Some of the bones show morphological overlap between *Australopithecus* and *Homo* species, which makes it difficult to distinguish them and to assign them to the correct genus (Johanson and Edgar 1996, 122).

The difficulty of distinguishing postcrania and assigning them to a species is one of the reasons for the choice of fossils included in this study. Instead of focussing on a small group of taxonomically strictly defined fossils, a large range is included. This increases the number of fossils and shows the size range present in this time range, which facilitates the making of comparisons between fossils and their body size estimates. Only very few fossils will be under close scrunity, they are used to study the trajectories of the estimation equations in order to assess which of these equations would be best used for *Homo erectus* stature estimations. The fossils that will be studies are not so much chosen on their taxonomic label, but on their size.

2.4 Discussion and conclusion

As we have seen in the previous chapter, information on body size can be very useful when one is drawing a picture of the life of early hominin individuals and the evolution of their lineage. The factors that influence the estimation of the body size of hominin individuals have been discussed in this chapter.

Stature and body mass are both useful measures of body size in hominins. Stature is considered to be more reliably estimated than body mass by some (Feldesman *et al.* 1990), but not by others (McHenry 1974). Stature estimation has as its advantage that there are more, and more reliable, physical anthropological data and methods and that femur or tibia length constitutes a large part of actual body length. However, McHenry (1974) states that the differences in body proportions between early hominins and the modern reference samples render stature estimations less reliable and useful than thought. As the exact nature of these differences is at present still unknown it is hardly possible to adjust for them in calculations.

The advantage that the estimation of mass has is its comparability to a broad spectrum of primate species, where it is mass that is measured and not so much stature. It is also more useful for energetic calculations that contribute importantly to the debate on *Homo erectus* lifestyle (Aiello and Key 2002). However, many of the human osteological collections have quite unreliable data on the body mass of the subjects (Steudel 1980), which would make estimation formulae that are based on them unreliable as well. Body mass can vary more than stature in the individual's life. The relationship of the predictor element with size is less direct (even when functional), as the weight of the bone itself does not contribute to the total mass as femur or tibia length does to total length. However, correlations coefficients still can be (very) high, and the estimations should, at least in the reference sample, be considered reliable.

The similarities between the reference sample and the species to be estimated are of paramount importance to the reliability and biological reality of the estimated body size. After all, a certain biological relationship is measured and applied to another population in which this relationship is unknown. Unfortunately, there are no perfect fits for extinct species. An extra difficulty in the hominin lineage is the existence of only one living representative, which is also the only species in the entire order of primates that shares the mode of locomotion with the extinct hominins. *Homo sapiens* populations are thus often used as reference samples for early hominin size estimation. This is somewhat better defensible for the *Homo* species than for the earlier australopithecine species, as there is evidence that (some) of the *Homo* individuals are quite similar in size and body proportions to *Homo sapiens* (Ruff and Walker 1993).

However, applying a regression equation made on one modern human population to another such population already gives deviant results, as was known even to early workers (Pearson 1899; Stevenson 1929). Applying a regression equation made on a modern human population to a different species, even if it is more similar to us than its predecessors, is not guaranteed to give accurate results. There are unfortunately few other methods that can be applied in its place. Workers have therefore focused on selecting the best-fitting populations (Kurki *et al.* 2010; Ruff and Walker 1993) or on taking a broader view (Feldesman *et al.* 1990; Hartwig-Scherer 1993; Jungers 1990; McHenry 1992).

From the discussion above on the different factors of size estimation that influence the results, several requirements regarding the data arose. These can be used to evaluate the methods. A predictor variable is generally considered to be useful when it has a high correlation coefficient, meaning a value above 0.8. However, a factor even more important in the choice for a certain predictor variable is its availability in the fossil record. Fortunately, the skeletal elements that are best preserved are heavy, large ones like the femur. The femur has elements highly correlated with body size, due to its function in carrying body weight and in locomotion. Other skeletal elements that are not functionally related to body size but do have a high correlation coefficient with it are used for size estimation as well. Cranial predictors have good availability and high correlation coefficients, but are not functionally related to body size. They might be very useful for estimating body size, but only if the observed high correlation coefficients in related species are present in the early hominin species. Since we have no way of knowing if this is true, it is the more reliable functionally related postcranial bones that have received the most attention as body size predictors.

Several paths are available for drawing up a body size estimate for an individual that is represented by a single fossil bone. All have their specific problems, but the largest problem is that there is hardly any possibility to test which of the methods comes closest to the past reality; there is one possibility for testing, but this can be applied only if a complete skeleton is found. Since this is rarely the case, other methods are sought for testing the estimation techniques.

Chapter 3

Material and Methods

The answer to the question of early hominin body size depends on the methods used to calculate it. The importance of the question and the uncertainty of the level of correctness of the answers, have urged researchers to search for different methods of estimation. The reference sample and statistical method, have been varied in a search for the best predictor equations. The results of these different methods, as expected, vary, even when used on the same fossils. When the goal is to establish the body size of a hominin individual, it is therefore necessary to make a choice from the available estimations and estimation methods. To aid in this choice, as a part of this thesis, I have collected and presented an overview firstly of the different estimation techniques and the variables they use, and secondly the results per fossil that these methods give. Using these data, I will try to argue which equations are more reliable than others. I will do this by weighing the different factors and equations in terms of their reliability and suitability. The collected data and the methods used to analyse them will be explained in this chapter.

3.1 Material

The material used in this thesis consists of stature estimation equations based on femoral measurements, which have been gathered from a number of publications (see the appendix). Use was made of forensic as well as palaeoanthropological literature. A total of 110 equations that use femur length as the only variable to estimate stature were extracted from 32 publications (see appendix). An effort was made to include as many equations as possible to give a complete overview. Information concerning these equations was also collected, such as data on the reference sample that has been used.

The equations are based on populations around the world, though the research intensity has been varied. Europe (Bach 1965; Boldsen 1984; Breitinger 1937; Eliakis *et al.* 1966; Hauser *et al.* 2005; Lorke *et al.* 1953; Maijanen and Niskanen 2010; Mollison 1911; Olivier 1963; Olivier *et al.* 1978; Olivier and Tissier 1975; Pearson 1899; Radoinova *et al.* 2002; Ross and Konigsberg 2002; Rother 1971; Telkkä 1950) and the USA (Dupertuis and Hadden 1951; Hens *et al.* 2000; Ousley 1995; Trotter and Gleser 1952, 1958; Wilson *et al.* 2010) have been seen a large share of the attention, while information on African (Allbrook 1961; Didia *et al.* 2009; Feldesman and Lundy 1988; Lundy and Feldesman 1987; Raxter *et al.* 2008), Pygmy African (Hens *et al.* 2000; Jungers 1988; Olivier 1976), Asian (Pan 1924; Stevenson 1929), indigenous American (Béguelin 2009; Genoves 1967; Ross and Manneschi 2011; Sciulli and Giesen 1993) and Australian populations has been much more scarce. Some of these publications could furthermore not been used in the current study, for example due to the use of predictor variables other than the femur or unavailability of the publication (Allbrook 1961; Didia *et al.* 2009; Radoinova *et al.* 2002). Some were unavailable but had their equations republished in other publications, e.g. (Ousley 1995 in Wilson *et al.* 2010) and (Bach 1965; Breitinger 1937; Mollison 1911; Rother 1971 in Wurm and Leimeister 1986). Some were published in a format that could not be easily transformed into an estimation equation in centimetres (Pan 1924 in Wurm and Leimeister 1986). The list of all equations is presented in the appendix.

Data on the lengths of 25 Pleistocene hominin fossil femora were collected from the literature, and applied to the stature estimation equations in Microsoft Excel. This resulted in a stature estimate per stature estimation equation for every fossil. These results are gathered in the same table containing the estimation equations and their associated data in the appendix. This allows for the quick association of a fossil, its stature estimate for a certain estimation equation and, e.g., the reference population it is based on.

Since the fossils are in effect random points in a predictor-input range of a certain width, a standardized dataset was also created by calculating the result per equation over a range (10–60 cm) of input variables (see the appendix, second sheet). This allows to show the behaviour of the equations over a larger input range and in a more standardized manner than when only the fossils are focussed upon.

These 110 estimation equations are not all independent of each other; many come in pairs of two, consisting of a male and a female equation for the same population. Of these 110 equations, 27 are based exclusively on females, 38 on males, and 14 on individuals from other species. The remaining 31 equations are based on both female and male humans. Whenever a selection of equations is necessary, it will be the male-only based equations that will be used because this is the largest unmixed group.

After compilation of the initial dataset, it transpired that some of the formulae from different publications used the same reference sample set. Since including all these formulae would inflate the influence of that dataset, some of them were removed in the course of creating the dataset that was to form the basis for the analyses. To provide an overview, the initial dataset is presented in the digital appendix. The total number of unique reference populations is 42, of which 40 are human reference groups and 2 are nonhuman hominoid reference groups. This last group is from now on referred to as the Hominoid group, which includes all nonhuman ape samples. The species included in this group are *Gorilla gorilla*, *Pan troglodytes* and *Pan paniscus*.

However, some formulae did use the same dataset while varying other factors, such as the mathematical method that they used. As it is of interest here to see the behaviour of the formulae when only one factor, e.g. the mathematical method, is varied, these are included. When other factors are varied, as for example separate formulae for the side of the bone, only one representative formula is chosen to eliminate noise and over representation.

In the case of the side of the bone, the formulae based on the right side have the preference. Similarly, when both maximum and bicondylar femur based equations are presented, it is the maximum femur that has preference. This choice has been made because when only one of these two femur measurements is used, it is most commonly the maximum femur length. A choice for maximum femur length thus increases the homogeneity of the sample. In some cases, more than two equations are based on the same reference group, as for instance when sex-specific equations are made separately on left and right femora. In such cases, the selection of one representative sample is made following the same guidelines as indicated above; in this case, the male-only and right-side based formula.

A wide selection is thus created that incorporates the entire dataset, and a strict selection that varies with the hypothesis to be tested. The strict selection has as its first goal to include only one representative equation per reference sample. Excluded are: all female-based equations, all RMA- and MA- based equations when OLS was available, and all combined-sex equations when a male only-based equation was available. The basic strict selection is indicated in the dataset (appendix) in the column “strict selection”. It has been further restricted for certain tests that were executed (this is indicated per test in Chapter 5). Excluded from both the wide selection and the strict selection were all equations that calculate skeleton height instead of living stature; adjusted formulae were made for calculating living stature using the published formulae for calculating skeleton height. The skeleton height calculating formulae are presented in the dataset, but are not counted within the equation counts.

This restrictive selection is made to exclude all equations possibly creating noise because they differed in other factors too (e.g. mathematical method), and to minimize the extra weighing of the reference samples on which many equations were based. The selection varies according to the hypothesis that is to be tested. This means that sometimes other equations are selected or excluded, for example, when only human-based equations are to be tested, all nonhuman ape-based equations are excluded. This is indicated per test in Chapter 5.

The information that was gathered with regard to the equations consists of the following variables for the reference sample: population provenance (nationality/ nationality + geographic origin); sexes used; total number of individuals used; range and/or average of femur length; range and/or average of stature; socio-economic background of subjects; state of the body when measured for stature (in life/ on corpse/ on skeleton (Fully's method)); state of the bone when measured (dry/ fresh/ in life (limb measured); and the correlation coefficient of the equation. This information will be used in evaluating the dependability of the equations. However, not all this information was available for all of the equations. When a selection of equations needs to be made, the equations that provide the above information are preferred over those that do not.

The femur is the preferential bone for body size estimation (e.g. Hauser *et al.* (2005)). Femur length can be measured either as maximum length or as bicondylar (physiological) length, see Jantz *et al.* (1994). The estimation equations use reference samples of which the femur is measured either as maximum or as bicondylar length. Even though it is more precise to use the corresponding measurement, usually only one of them is given for a certain fossil, and Feldesman and Lundy (1988) conclude that the difference between maximum and bicondylar length is so small that it hardly affects the results of the calculated body size. Therefore, the measurements given in the literature will be used if they are either maximum or bicondylar length on all formulae, though maximum length will be preferred.

3.1.1 Fossil femora

Data concerning the lengths of fossil femora form the raw material for this study, and although the discussion in this thesis primarily focuses on *Homo erectus* size estimation, hominin femora from a large size range have been included. This allows for a better study of the behaviour of the estimation equations.

The fossils incorporated in this study have been selected primarily on the basis of the opportunities they provide for measuring or calculating the length of the individual's femur. This includes complete femora, parts of fossil femora and also os coxae in which the acetabulum is present. The age of the fossils included is 2 my or younger, with a focus on fossils regarded as (possibly) *Homo*, especially *Homo erectus*. No geographical restriction has been placed on the fossils.

The femoral lengths of the fossil femora were derived from several methods. Firstly, some of the femoral lengths come from the measurement of (semi-) complete femora (e.g. D4167, Lordkipanidze *et al.* (2007)). However, many fossils are incompletely preserved. They can be reconstructed used other, similar-sized and more complete fossils or in the cases where a femoral head is preserved, they can be predicted using a regression equation. For this purpose, the equation provided in McHenry (1991a) is used, which is also the source for many of the fossils' femoral lengths and ages. For fossil femur lengths, the numbers provided by McHenry were used because of his extensive work on the reconstruction of femoral lengths, with the exception of the femur head estimation equation for femur length estimation, which I executed myself. These data from McHenry (1991a) are supplemented with information from Simpson *et al.* (2008) (Gona), Feldesman and Lundy (1988) (KNM-ER 999), Lordkipanidze *et al.* (2007) (D4167), Conroy (2005) (Trinil, Zhoukoudian) and Grine *et al.* (1995) (Berg Aukas).

Pelvic fossils have also been included, as femoral head size can be estimated from acetabular size and can subsequently be used in the femur length estimation equation proposed by McHenry (1991a). Femur head estimation formulae and other methods are presented in Simpson *et al.* (2008).

The data that are used here are the single estimated statures; no confidence intervals were added. In a way, this does not do justice to the equations, as in almost all publications the uncertainty of the results and the need for carefulness are being stressed. However, in the analysis of so many equations at once, the addition of confidence intervals would obscure the behaviour that is under study. In actual stature estimation of a single fossil, confidence intervals and error should be considered.

The fossil that is used as input in a estimation equations determines the result. The choice has been made to use OH 28 as input in all tests. This fossil has been chosen because of its femur length, 45.6 cm, which is close to the femur length mean in many modern human populations (see appendix). In addition, some tests have been executed on the smallest fossil in the sample (KNM-ER 1463, 31 cm), on the largest fossil in the sample (Berg Aukas, 51.8 cm) and on D4167 (38.6 cm) because of its position in between the values of KNM-ER 1463 and OH 28. The entire range of the fossil femoral lengths is thus represented by these four discrete input values.

3.2 Method

3.2.1 Selecting the best equation: different approaches

The central aim of this thesis is to find the best-fitting and most reliable stature estimation method for *Homo erectus*, based on femur length. As a first step, an overview of the available equations has been created. A three-fold approach will be taken to assess these equations with respect to reliability and goodness of fit. The first approach is to assess the gathered equations in terms of a priori quality criteria (Chapter 4). The second approach is to take a closer look at the behaviour of these formulae and to test hypotheses regarding grouping criteria (Chapter 5). These approaches are followed by test-cases of three hominin femur fossils which will be used as examples of the practice of stature estimation and serve as an analysis of congruence of different estimation techniques (Chapter 6).

The first approach, an a priori quality criteria assessment, will use the criteria defined in Chapter 2. The factors that constitute an equation were found to be predictor variable, reference sample and mathematical method used. For this thesis, the choice has been made to concentrate on femur length as the predictor variable. Femur length is generally thought to be the most reliable single predictor for stature (Damuth and MacFadden 1990) and femur fossils are well-represented amongst hominin fossils.

The other factors do vary. No equations were excluded on the basis of their mathematical method or reference sample. On the contrary, an effort was made to include a diverse set of mathematical methods and reference samples. But where reference samples from all over the world have been used in stature estimation equations, the diversity of the mathematical methods involved is more limited.

The criteria to evaluate these factors have been discussed in Chapter 2. Whereas criteria for the mathematical method used were largely undetermined from that discussion, a large set of criteria for reference samples did emerge. Therefore, reference samples can be evaluated with this set of criteria from the literature, but mathematical method can not. These criteria for reference samples are the correlation coefficient, the N and condition in which stature and bone length were measured.

In addition to the evaluation of factors on a priori quality criteria, evaluation of equations can be executed by the comparison of their results. This is the second approach. It enables the identification of the patterns that lie at the base of the differences in results of the estimation equations. Hypotheses have been proposed concerning these underlying reasons of variation in estimation result between populations. These are, for example, climatic adaptation (Ruff 1991) and ethnic origin. To test these hypotheses, the equations and their results will be grouped according to the different subgroups present, for example, by ethnic origin groups. ANOVA will be executed to test for significant differences between these subgroups. If significant differences between subgroups are found, this is an indication of the validity of these groups and therefore their influence on stature estimates.

The results are explicitly analysed per fossil and not per fossil taxon. Patterns in time can be obscured by grouping fossils under a taxon, which obviously has a much longer “running time” than a single individual (Robson and Wood 2008). The taxonomic debate shows that it is difficult to pinpoint almost any fossil other than the type specimen to a particular species. The postcrania which are most frequently used in these body size estimates, are even more notoriously difficult to assign to a species. It is therefore not attempted or considered appropriate to extend any conclusions from one fossil to another in this thesis, as no further taxonomic research is executed. It is not attempted to come to estimates for all fossils included in the dataset or to come to a species average for *Homo erectus* due to limitations of the scope.

Significance levels are given for $p = .05$, though it is indicated when a significance for $p = .1$ is attained in the overview of results. In the text however, significance of $p = .05$ is simply indicated with an asterisk following the p-value. One standard deviation as calculated by IBM SPSS Statistics 19 is indicated with ± following a result.

3.2.2 Reference sample data quality evaluation

The several factors that are involved in the estimation technique have been discussed in Chapter 2. The requirements that the data of these factors should meet can be used as points of evaluation of the estimation technique. If the requirements are met, the estimation technique will in general be more reliable than is the case when they are not met. In Chapter 2 these requirements have been summarized as in the list below. These criteria will be used for evaluation of the reference sample data in Chapter 4.

- Measurements: bone or limb segment; stature (in vivo/on corpse/Fully's)
- Reference sample: sample size; reliable data; associated data; age distribution
- Fossil match: average body size; average femur size

3.2.3 Estimation equation result evaluation

The calculated results from the estimation techniques will be evaluated using boxplot figures and estimation techniques trajectory graphs. Boxplot figures will give an overview of the total range of predicted statures per fossil. Outlier and clustering equations can be identified in this way. These findings can be combined with information about the method (e.g. reference sample used) to explain the reason for the patterns. The estimation techniques trajectory graphs can aid in understanding these patterns.

The reference sample groups will be pooled into larger "reference sample groups". They will be subjected to T-test and ANOVA tests in IBM SPSS Statistics 19, to test for differences between the results from each of these groups. In this way, the influence of the grouping characteristic on the final estimation result will be evaluated. Levene's test (for equality of variances of the different groups tested) was checked for all T-test and ANOVA runs. When it is not mentioned, Levene's test was insignificant. If Levene's test did give significant results, a Welch test was executed as is recommended when the homogeneity of variance assumption of ANOVA is broken (Field 2009, 379).

When these tests give significant results for certain groupings, e.g. on ethnic origin, this is an indication that that classification groups according to biologically relevant criteria for stature estimation. This then indicates that this classification offers relevant criteria on which a reference sample and an extinct homonin should be matched, in order to reduce the estimation error.

The reliability of the techniques will be assessed by determining the level of agreement between their results and those of other techniques that are based on different data. The accuracy of the estimation equations can be tested by using several methods based on different principles and applying all of these to the same hominin individual, in the manner done by Geissmann (1986a). This way, the results of different methods such as regression analysis, anatomical stature and/or techniques based on completely different reference samples (e.g. species) can be compared. Another possibility is to examine the agreement between methods based on different bones from the same individual, e.g. femur, tibia and/or cranial variables. These methods will be used in Chapter 6 on several fossils, as test-cases.

3.2.4 Estimation equation classification hypotheses

The first selection to be tested are the male-only based equations and the female-only based equations. These are compared with an independent samples T-test on the means of male and female equation results (test 1.1, see table 3.1). The equation selection used in this test is formed by the reference samples for which both a female and a male equation are available.

The second selection will test the effect of the mathematical method used by ANOVA on OLS, RMA and MA (test 2.1, see table 3.1). Excluded from the equation selection are all articles that do not explicitly test differences in mathematical method. Included are the data from Feldesman and Lundy (1988); Hens *et al.* (2000); Jungers (1988); Maijanen and Niskanen (2010). For the second test (test 2.2) of this selection, the log-transformed equations were separated from the raw equations, and only the latter were tested in the ANOVA. The log-transformed groups were not included in an ANOVA test as separate groups as they contained only 2 equations each, which is a very low N.

Table 3.1: The tests executed on the reference sample hypotheses

Tests	**Tested by**	**Groups**	**Selection**	**Fossil**
1.1	T-test	females and males	groups with female and male equations	all
2.1	ANOVA	OLS, RMA, MA	groups with OLS, RMA and MA equations	OH 28
2.2	ANOVA	OLS, RMA, MA	idem, excluding logged equations	OH 28
3.1.1	T-test	nonhuman apes, humans	wide	OH 28
3.1.2	T-test	nonhuman apes, humans	raw-only	OH 28
3.1.3	T-test	nonhuman apes, humans	strict	all
3.2.1	ANOVA	Europeans, Africans, Asians, Americans	strict	all
3.2.2	T-test	Europeans, Africans	wide	all
3.2.3	T-test	Europeans, Africans	strict	all
3.3.1	ANOVA	Af. Am., Cau. Am., Ind. Am., mixed Am., Europeans, Af. Pygmies	strict	all
3.3.2	ANOVA	Af. Am., Cau. Am., Ind. Am., Europeans, Af. Pygmies	strict	all
3.3.3	ANOVA	Af. Am., Cau. Am., Ind. Am., Europeans	strict	all
3.3.4	ANOVA	Af. Am., Cau. Am., Ind. Am., Europeans	wide	all
3.3.5	T-test	Af. and Cau. Am.	strict	all
3.3.6	T-test	Af. and Cau. Am.	wide	all
3.4.1	ANOVA	size groups	wide	all

Reference sample influence

In the literature on stature estimation in hominins it has often been proposed that the reference sample on which an estimation equation is based is of paramount importance to the final result, with different reference samples giving different results for the same fossil. In modern humans, estimation are preferentially done by using an equation from the population to which they belonged. This gives more accurate results (Hauser *et al.* 2005; Pearson 1899; Stevenson 1929; Trotter and Gleser 1958). The reason for this greater accuracy is not entirely clear. It has been proposed that it is due to similarities in body proportions that are caused by genetic relatedness. Another possible cause of such similarities is the fact that these individuals all live in a similar environment to which they supposedly adapt during growth or are genetically adapted to because their ancestors lived there too.

These are the two main hypotheses that explain differences in estimation results from differences in the populations on which the equations are based. The first hypothesis, then, groups populations according to"race", for example in the groups African, Caucasian, Asian as do Feldesman and Fountain (1996). The second postulates that populations should be grouped according to the environment (climate, altitude, latitude) that they inhabit, as their body proportions are hypothesized to be adjusted to this (Ruff 1991, 2002; Ruff and Walker 1993).

If either of these hypotheses are true, it can be inferred that reference samples do influence the results of estimation equations based on them. To test the genetic relatedness hypothesis, reference samples will be grouped according to assumed genetic closeness. Due to the poor availability of estimation equations, this will unfortunately be done in groups covering large geographical areas. The climate hypothesis can not be tested because the data needed for this kind of testing are not available.

Several tests were designed to test different classifications of the origin of modern human populations. These classifications all follow the ethnic origin hypothesis. No attention was given to other differences between the groups, like climatic similarities. The first test is a baseline test, distinguishing humans and nonhuman hominoids. If the results of these groups can not be separated, it would be expected that no human groups can be distinguished from each other as all humans are more similar to each other than to any of the hominoids. The human and nonhuman samples were compared with each other in an independent samples T-test (test 3.1.1, see table 3.1). The first test is executed on a wide selection of equations, but excludes the combined human/nonhuman equations. The second test (test 3.1.2) is executed on the same selection but excludes the log-transformed equations. The third test (3.1.3) is executed on the strict selection (i.e. only one equation per sample, no log).

The second test follows the classification of Feldesman and Fountain (1996) with some adjustments. Feldesman and Fountain (1996) divide all modern humans into three racial groups; Caucasian, African and Asian in origin. The classification that was followed for the test was that on the continent of origin of the populations, separating the groups from Europe, Africa, Asia and the Americas. Such a division should test if these continent-scale populations are different from each other in estimation results (Feldesman and Fountain 1996). If so this is presumably caused by large-scale genetic ties, as the populations representing the continents live in mixed environments and climates (see appendix on reference sample provenance data). The first test (test 3.2.1) for this classification was an ANOVA on the strict selection. T-tests on the European and African group using the wide and strict selection followed.

In other publications (Feldesman and Fountain 1996) the group "Indigenous Americans" is included within the Asian group. However, it has become clear that Indigenous American groups appear to have body proportions that are quite different from those of Asians (Sciulli and Giesen 1993). Furthermore, it is reasonable to say that Americans lived in isolation from the Old World populations for a very long time, an isolation that began perhaps shortly after the initial colonization of the Americas, and so probably have been developing independently for some 10,000 years.

It is recognised that the continent-of-origin classification might be quite rough, and has some unwanted groupings. For example, African Americans were grouped with the African origin group, while it is known that they can be as different from current Africans as they are different from Caucasian Americans in tibia/stature and femur/stature proportions (Ruff and Walker 1993, 246). The third classification is therefore based on finer grained geographic provenance criteria, resulting in the groups listed below. Pygmies were indicated as a separate group, because they are treated as such in the literature (e.g. Hens *et al.* 2000; Jungers 1988; Olivier 1976). Of these groups, 6 were used in an ANOVA test on the strict selection (test 3.3.1). These groups are indicated with an asterisk in the list below. A second test is executed on the same conditions, but eliminates the Mixed American group. The third test idem dito, but without the Pygmy sample. The same groups but with the wide selection of equations are tested on a fourth ANOVA. Test 5 is a T-test comparing African and Caucasian Americans in a strict selection. Test 6 is the same T-test but with a wide selection.

1. African Americans*
2. Caucasian Americans*
3. Asian origin: Americans (mixed Asian) and Asians
4. Indigenous Americans* (both North and South Americans, including Inuit)
5. Mixed Americans* (also mixed Indigenous American - Caucasian)
6. Europeans*

7. Africans (sub-Saharan)
8. African Pygmies*
9. Other human
10. World-wide humans
11. Nonhuman ape
12. Mixed human-nonhuman samples

A final classification was based on the average stature of the reference samples. The dataset was divided into four size groups, which are listed below. No restrictions were placed at the equation selection.

1. <150 cm on average
2. 150 – 159.9 cm
3. 160 – 170 cm
4. >170 cm

3.3 Conclusion

Reliability can be evaluated by comparing the results of the different methods. The reliability and accuracy of the techniques brought together in this study can be assessed using the criteria discussed above. Which of these techniques employ methods or samples that are expected to yield less reliable results? And which of the results from different techniques agree with each other, thus adding to the probability that these estimations are indeed accurate? In the next chapters, the results of the three approaches that were explained in this chapter will be presented.

Chapter 4

Reference sample quality

In this and the following two chapters the results of several approaches to elucidate the mechanisms of body stature estimation will be presented. Three approaches will be taken. The first will present a closer look on the quality of the equations, by evaluating the information on which they were built, see section 4.1. This approach will be attended to in this chapter. The criteria that were identified in Chapter 2 will be shortly brought to mind again after which these criteria are used to evaluate the reference samples and their equations in the dataset. These results are subsequently discussed to yield useful information to aid the choice of the equations and/or sample best used for *Homo erectus* stature estimation.

The second approach is the most voluminous section of the results, and presents the results for several statistical tests executed to identify the most influential factor in body stature estimation. This is presented in Chapter 5. The third approach is a test-case in which information gathered from the first two approaches is used to explore body stature estimation in specific fossils. This will be in Chapter 6.

4.1 Criteria

The equations presented in the previous chapter will be evaluated according to several criteria. The techniques can be dissected into their factors and compared to the recommendations discussed in Chapter 2. The factors which an estimation technique consists of have been discussed in Chapter 2 to be predictor variable, reference sample, and mathematical technique. The estimation techniques presented here have been selected for their predictor variable (femur length) and mainly differ in the choice of reference sample. Therefore, most criteria used here evaluate the reference sample used.

There are certain requirements that, when fulfilled, would make a reference sample more reliable for body size estimation. In this section, the estimation equations used will be discussed following these criteria. These criteria can be formulated negatively as avoiding the following indications of lower reliability of an estimation technique.

- A low correlation coefficient, for example by the use of predictors that have a low correlation with body size; e.g. upper limb instead of lower limb bones (Trotter and Gleser 1958), or living limb segment measurements instead of bone measurements.
- Small reference sample sizes, a minimum of 30 to 50 individuals is taken as a standard. Pearson prefers even more: *"When the correlations of the organs under consideration are high (e.g. long bones in Man), fifty to a hundred individuals may be sufficient; in other cases it is desirable that several hundred at least should be measured"* (Pearson 1899, 170).
- Unreliable data for the reference sample, e.g. unclear measurement sets, body size measurements recorded long before death or taken from statements of relatives.
- Unassociated body size and bone measurements (not from the same person).
- Unrepresentative samples, e.g. only old, diseased or individuals of a certain social class are included.
- Unclear population history and/or large-scale mixing of populations makes it more difficult to match such a population to a hominin individual in terms of, e.g., geography.
- Averaging equations from different populations and sexes.
- Extrapolation necessary due to differences in size between the reference sample and goal individual.
- Compound error by estimating femur length from a femur fragment and from that body size, both by regression.

This can be summarized as evaluating the following characteristics:

- Measurements: bone or limb segment; stature (in vivo/on corpse/Fully's)
- Reference sample: sample size; correlation coefficient; [reliable data; associated data; age distribution]
- Fossil match: [average body size;] average femur size

The information per estimation equation is presented in the table in the appendix and will be discussed point by point here.

4.2 Results

4.2.1 Measurement type

Bone measurement is considered better than limb measurement. In limb length data collected on live subjects it is the true measurement of the femur that needs to be estimated, thereby potentially introducing error. Publications that use live subjects are Breitinger (1937); Olivier (1976); Olivier *et al.* (1978); Olivier and Tissier (1975). Hauser *et al.* (2005) measured bones in a fresh state, with cartilage still attached, because he has a forensic aim. In all other situations when the measured state was known, the bones were measured in dry state.

In stature estimation it is living stature that is the goal of the estimation. The best stature measurement in the reference sample would therefore be during life. Measurement on the corpse, or on the skeleton in supine buried position, would be second-best, while measurement of the individual bones of the skeleton followed by a correction factor would be the worst of these methods because it has the largest potential for error. Reported stature from relatives or official accounts also have a large error margin (see Wilson *et al.* 2010).

Most of the reference sample subjects have been measured as a corpse. This probably only introduces moderate error (Hauser *et al.* 2005). Some of the authors recommend adjusting the calculated stature by 2 cm (Pearson 1899). A reported stature from information of relatives or documents is used in several cases. A reference sample of Olivier (1963) and Lorke *et al.* (1953) consists of prisoners of WWII, Wilson *et al.* (2010) uses American forensic cases, and Hens *et al.* (2000) do likewise.

Fully's method is used in a substantial amount of reference samples, in its original format by Fully (1956) or the adjusted version of Raxter *et al.* (2006). It is often applied in the case of old skeletal collections (Feldesman and Lundy 1988; Jungers 1988) as reported stature might be unreliable, and in the case of archaeological samples (Maijanen and Niskanen 2010; Sciulli and Giesen 1993). These reference samples are of special interest as they are of populations that are in all probability unmixed on a continental scale and extend the stature range. Although their living stature might thus have been measured with more error than the samples which have been measured in vivo, they have other useful characteristics that make up for this.

Some of the reference samples stand out in having in vivo measurements made shortly before death combined with bone (and not limb) measurements; the circumstances that enable this are those of war (Trotter and Gleser 1952, 1958). Reference samples that have limb measurements are measured in vivo for stature as well.

4.2.2 Reference sample

Seriously low sample size, below 20 individuals, is limited to 6 groups. Unfortunately but understandably, these include some of the special groups: two are Pygmy-based, two are bonobo-based and one includes Inuit people. The last group is a regular modern human group, Polish women.

The correlation coefficient between predictor and stature is roughly judged to be good when above $r = .8$. The greater part of the reference samples have a good r according to this rule-of-thumb. Notable exceptions to this are some of the hominoid samples, several samples measured in vivo and an US African American military sample. The hominoid sample is composed of several species, so a low correlation coefficient in comparison with the single-species groups is not surprising. Likewise unsurprising is the observation of a lower r amongst the samples of which the limbs have been measured instead of bones. This procedure includes an extra regression step (with an r of $> .9$) which would make the final correlation between the calculated femur length and stature lower.

Strikingly high r values ($> .9$) are attained in some samples, which all share Fully's method as stature measurement method. An explanation for this high correlation would be that the stature is in part based on the femur length measurement and the variables are thus not independent. The r in these cases does not have value as quality criterion.

The other factors that could influence the reliability of the equations (reliable data, associated data, age/class distribution) are not evaluated here separately, because quantified data on a large enough part of the dataset are lacking. They are addressed in the discussion on this subject below.

4.2.3 Fossil match

The aspects of stature range or femur size range are not to be judged according to a priori criteria like the aspects treated above. Their appropriateness for a hominin individual is depended on the size characteristics of that hominin individual, and should therefore be individually matched. Some general remarks can be made.

The femur range of fossils is larger than that in modern humans. In the dataset compiled here, fossil femur lengths range from 30 cm (KNM-ER 1463) to 51.8 cm (Berg Aukas). The compiled modern human reference sample femur length averages are typically between 42 and 49 cm. There are a few notable exceptions at the lower side. Pygmies have a low femur length average of, e.g. 37.7 cm. (Hens *et al.* 2000). Some indigenous Mesoamericans also have a low femur length average, of 39.6 cm. (Genoves 1967). The range within samples is substantial; the minimum of the samples with reported femur length range is between 34 and 40 cm, while the maximum is between 47 and 54 cm. It is for this reason that modern human groups with deviant femur sizes and other hominoids are included, as they extend the femur range which might be useful to gain insight in hominin body proportions.

4.3 Discussion

Based on this view on the dataset, it appears that on average reference samples hold up to elementary criteria. They consist of large enough samples and try to minimize potential sources of error by measuring the characteristics in such a condition that it approaches the state of the goal individual. In this section I will discuss the features of the samples in more detail and view them from a hominin estimation perspective.

One observance that stands out in the comparison of this set of reference samples, is that there are more reference samples based solely on men (38) than on women (27). This might be an effect of sexism in science, where the natural research focus was on men rather than on women. Most of the men-only publications are pre-1980 (e.g. Breitinger 1937; Lorke *et al.* 1953; Olivier 1963; Stevenson 1929). Furthermore, equations based on women are in total of lesser quality than the male-based equations. Trotter and Gleser (1952, 1958) stand out in quality among the male-based equations due to their reference samples and measurements, but use a less-than-ideal reference sample for their female equations. There also appears to be a trend that female-based equations are based on a lower number of individuals than the male-based equation from the same publications (e.g. Dayal *et al.* 2008; Feldesman and Lundy 1988; Genoves 1967; Hauser *et al.* 2005; Raxter *et al.* 2008; Sciulli and Giesen 1993; Telkkä 1950; Wilson *et al.* 2010).

There are also reference samples in which the sexes are grouped (e.g. Béguelin 2009; Hens *et al.* 2000; Jungers 1988). As it is expected that females and males from the same population give different estimation equations, this is not ideal because the effect is the mixing of two signals. Two of the samples involved however, are valuable because they represent rare reference samples: prehistoric Argentinans and Pygmies. Equations based on mixed sex will therefore not be excluded from the analysis in Chapter 5. When besides a mixed equations, also a sex-specific equations is offered (Hauser *et al.* 2005; Maijanen and Niskanen 2010), it is the latter one that is used.

4.3.1 Sample size and research focus

The early samples used for stature estimation (e.g. Pearson 1899; Stevenson 1929; Telkkä 1950), but also later ones (e.g. Béguelin 2009; Hauser *et al.* 2005; Maijanen and Niskanen 2010) were quite but not extremely small. They were measured after death as corpse or skeleton, but with the goal for stature estimation which made the investigators aware of stature change due to death. Some of them have proposed correction factors for converting corpse length into living length (Pearson 1899). These collections were collected with the purpose of body size estimation, which is not the (prioritized) goal of other, larger, skeleton collections. These collections, like the Terry and Hamann-Todd collection, were amassed in the first half of the 20th century in order to build a database of human skeletal variation. The Terry collection for example, consists of individuals from the state of Missouri (USA). The collected subjects until 1955 were individuals whose corpse fell to the state as nobody claimed their remains or because the family signed off the responsibility to the state. The demographic composition of the collection was due to this method of collection mainly of people from poor socio-economic background (Smithsonian Institute 2012). Their bodies are most likely influenced by their poor environment during growth, and may not be representative of people who grow up in an affluent, more healthy environment. After 1955 the demographic composition changed because of a change in the law, which required consent of the individual itself. A larger part consisted now of individuals from middle class background (ibidem). Thus while these collections present many more individuals (Terry collection: 279,804 (1967)), their background is mixed (and largely from poor socio-economic level), and body size records are not always present or reliable. A larger reference sample is thus not automatically a better reference sample.

4.3.2 The Trotter and Gleser samples

Equations that enjoy widespread use are those of Trotter and Gleser (1952, 1958) and Trotter (1970). They are popular because their reference samples meet many of the requirements. The sample consists of a relative large number of subjects with associated body size measurements during life and bone measurements after death, measured following a defined measurement set, measured in life within a few years of their death, and representing mostly young adults.

Such an exceptional reference sample opportunity resulted from several wars that the USA took part in, in which many of the soldiers were killed. The resulting estimation equations were separated according to “race”. Most of these military war victims were draft soldiers. They were called to the army without reference to their socio-economic background and thus will probably be a more representative cross-cut of the male young adult American population than the civilian collections like the mentioned Terry collection.

The downside of these Trotter and Gleser (Trotter and Gleser 1952, 1958) equations is that they can only be used on men. The female-based equations are from skeletal collections (Terry collection) which do not have the benefits that the military collections have. Some of the male-based equations have a low N. The two properly-sized groups are the African American group and the Caucasian American group. Furthermore, it seemed necessary to make adjustments on the tibia measurements due to apparent mismeasuring (Jantz *et al.* 1994). The formulae have been shown to overestimate stature on small-bodied individuals, not only australipithecines (Geissmann 1986a) but also indigenous South African individuals (Feldesman and Lundy 1988).

The military collection equations published by Trotter and Gleser (1952, 1958) have been very influential and much used in the estimation of stature of both modern humans as extinct hominins, for example on australopithecines (McHenry 1974), on *Homo erectus* by Leakey and Walker (1985) and Ruff (1988) and on Neanderthals by Trinkaus (1981). More recently, they were applied on the Dmanisi material (Lordkipanidze *et al.* 2007).

However, notwithstanding the excellence of the collections in some respects, it falls short on other aspects when to be used on extinct hominin species. First of all, the collections only comprise male individuals, which inhibits their use on modern females as well as presumably on extinct ones. Secondly, the subjects are modern North-Americans, which is likely a group with a higher average for stature then many other modern and extinct hominin groups (see for example the appendix). This raises the problem of extrapolation. They have also probably experienced conditions during growth which were very different from the conditions extinct hominins experienced. In conclusion, this collection might not be the most representative of the subgroups that our fossil individuals belong to. The second problem is also valid for the civilian American collections.

4.3.3 Small-bodied reference samples

Other reference collections have been proposed to overcome some of the mentioned problems. For expected small species like the australopithecines, modern small-bodied African reference samples with a hunter-gatherer lifestyle have been suggested (e.g. by Feldesman and Lundy 1988; Olivier 1976). Small-bodied reference samples in the size range of the prediction sample prevent extrapolation, which increases likely error.

A point brought especially to light by Olivier (1976) is the discrepancy between the average stature of the reference sample and the expected stature of certain extinct hominins like the australopithecines. These values are so far apart that extrapolation is required which will result in overestimation of the stature of the australopithecines. Oliver therefore created a reference sample of Pygmies from Cameroon (“Baka”). The average male stature of this group (91 individuals) was 154.5 cm, which is indeed a lot shorter than the average stature for the Trotter-Gleser samples (being 174 cm for the Caucausian Americans and 172.1 cm for the African Americans in Trotter and Gleser 1952). There is little skeletal data on Pygmies so Olivier uses limb segment lengths from living Pygmies to estimate limb bone lengths, which are used to estimate stature. This introduces compound error. Further, it appears that the correlation between femur length and stature is significantly lower in

Pygmies than in other modern humans (Feldesman and Lundy 1988, 586; Olivier 1976). It might be that this is an effect of using living limb lengths, because it is not seen in the other Pygmy samples (Hens *et al.* 2000; Jungers 1988). However, the stature measurement of these other Pygmies has been done via Fully's anatomical method which probably aids in attaining a high correlation coefficient thus rendering the correlation coefficient not very useful for quality comparisons.

Feldesman and Lundy (1988) aim to give an alternative for the commonly used stature regressions of Trotter and Gleser (1952, 1958) by using a different reference sample. Where the reference sample of Trotter and Gleser (1952, 1958) is of mixed North-American ancestry and social status while limited in age of the individuals, Feldesman and Lundy (1988) prefer to use a homogenous non-western reference sample which is a sample of several small-bodied South African (hunter-gatherer) populations of the Nguni, Sotho and Venda tribes (housed in the Raymond Dart Collection). Fully's anatomical method (Fully 1956) was used to determine skeletal height; this measure of stature was used in the equations. To reach living stature, Fully and Pineau (1960) recommend a correction factor to account for soft tissues (between 10–11.5 cm, depending on skeletal length). This is different to most other techniques, that regress on living stature and thus give a living stature result. The need to use Fully's (debated) correction factors increases the uncertainty of the result (Porter 2002).

4.4 Conclusion

A diversity of stature estimation equations are available for the calculation of a stature for the length of a given femur. How to choose one from among them? When the goal individual of the estimation is a recent modern human, it is common practice to use an equation based on a reference sample which comes from the same population, the same generation (birth cohort) and the same sex as the goal individual. As there are no estimation equations for every population, the "same population" guideline can get a bit stretched. For example, in the Netherlands the formula of Breitinger (1937) based on athletic male Germans in the 1930s are used in addition to the Trotter and Gleser (1952) formulae for Caucasian Americans (Maat and Mastwijk 2005). While the guideline of matching sex can be followed, the matching population guideline comes down to matching "race" and the generation guideline is completely ignored.

A quality evaluation of the available stature estimation was executed to deliver insight in the quality differences between the reference samples on which these equations are based. This can help in the choice of the best estimation equation. Several measures of quality were discussed in Chapter 2 and used in this chapter to evaluate the equations gathered. Most of the equations sufficed to the basic quality criteria that were checked here.

High quality data that were present with almost all samples on which this could be checked (i.e. data were reported in the publication of the equation) were, a.o., a high correlation coefficient between predictor variable and stature and the measurement of femur instead of the thigh. The first one, though vital for an accurate estimation, is unsurprising. It is due to the choice of predictor variable, which is the femur length in all gathered equations. The femur is known to have a very high correlation to stature, which is exactly the reason for its usage in stature estimation on such a scale.

Most reference samples also used the femur length itself, and not the femur length as calculated from the thigh length. This is the outcome of a dilemma of quality. Stature is more accurately measured during life while femur length is more accurately measured on the (dry) bone which requires the death of a subject. In most cases it has been settled with stature being measured on the corpse or on the skeleton. In some samples, skeletal material is hard to come by (e.g. Pygmies) and both stature and thigh are measured during life.

There appeared to be a difference between the reference samples that were purposefully gathered to function as a stature estimation sample and the general osteological collections. The second kind has a much higher number of individuals and would therefore on first thought considered to be qualitatively better, but the information of the individuals in such collections can be of lesser quality, something that cannot be made up by a larger N. The information that is concerned is, e.g., the accurate measurement of stature which is reliably associated with the skeleton. Purposefully collected samples have been given more attention to e.g. the change of stature due to death.

Some of these smaller collections might also reflect a more homogenous population, because they sample archaeological populations from before the advance of large-scale mixing (e.g. Maijanen and Niskanen 2010; Raxter *et al.* 2006; Sciulli and Giesen 1993). For the purpose of choosing the best reference sample for an estimation, it is of interest to collect reference samples that are homogenous and unmixed. This is to prevent several signals from the same sample.

Hominin individuals might require additional information next to general quality concerns. One of the few certainties when it comes to hominins is that there is a wider range of femur lengths than represented by the standard modern human samples. It is therefore useful to enlarge the total femur range by including small-bodied modern human samples in the dataset. This is the reason that even if it is not recommended to use samples that use limb segment measurements instead of bone measurements, the sample of Olivier (1976) will be used in this thesis because it is one of the two samples that is available. The two other publications that published a Pygmy-based equation are both based on the same sample (Hens *et al.* 2000; Jungers 1988), and consequently only one of them is used to avoid over representation of that specific sample. The benefit of using Pygmy samples is in this case considered stronger than the benefit of using samples that use femur measurements.

The equations which are deemed of good quality in this chapter are not necessarily the best equations for extinct hominins. The better the quality of the reference sample, the more likely it is that the estimation equation correctly represents the population. It will only give more accurate results for an extinct hominin if the extinct hominin had the same body proportions as the human reference group, so if this individual would fit in this human group. However, these good reference samples are more likely to give the signal of the population they represent and not a signal distorted by noise created by low N or other problems. This makes them more useful for the analysis in the next chapter

The preference for an equation stemming from the same population as the goal individual shows is based on the long-term knowledge that such an equation gives a more accurate result than does a random equation based on another population. This has been already shown by very early researchers in this topic (e.g. Pearson 1899). It can be assumed that this is because there is a shared element between individuals in a population. This could be due to shared environmental (climatic) circumstances, shared socio-economic (nutrition, health) circumstances, shared genes either by genetic drift/founder effect or by adaptation, or due to some other reason. If it would be known which influences the result of the stature estimation the most, a choice for a hominin individual would be better founded. Therefore, in the next chapter it will be attempted to assess which factors influence the stature estimation the most.

Chapter 5

Estimation equation patterns

This chapter concerns itself with the results of several statistical tests executed to identify the most influential factor in body stature estimation. It is separated in a section that presents the results for the three main hypotheses tested and a section that analyses these results. A discussion on the results is included.

The goal of the following statistical tests is identifying the causes of the differences in results of equation estimations. There are several general hypotheses on the causes of these differences.

A. Adaptation to climate by changing of body proportions (Ruff 1994).

B. Ethnic differences; different "races" have different anatomical characteristics, for example due to genetic drift.

C. Differences in average height between reference samples used.

D. Differences in statistical technique used (ratio; OSL, RMA, MA; log transformations).

E. Differences in results are due to random variation between human populations.

To test these general hypotheses, selections of equations that have a certain aspect in common (e.g. ethnic origin, statistical technique) will be made. When there is high within-group coherence combined with an inter-group significant statistical difference this is taken as evidence that the grouping factor does indeed play a role in causing the estimation result. When the cause of the differences can be determined, a choice of estimation equation fitting for extinct hominin species can be made on better grounds.

Not all of these hypotheses will be tested here. The detailed environmental data needed for the first hypothesis, climate adaptation, are not available for this study. Consequently, the focus will lie on the evaluation of points B, C and D; the influence of geographic/ethnic origin groupings, of average height of the samples and of the mathematical technique used, on the results of the estimation equations.

These general hypotheses have been reworked into several testable hypotheses containing subtests for statistical analysis. The testable hypotheses have been arranged into three sections. The first tests the difference between the means of female and male-based estimation results. This has not been included in the general hypotheses above, but is tested as an examination of the data before turning to testing the actual hypotheses. The second tests the differences between the means of equations calculated using OLS, RMA or MA mathematical methods. In the third section the influence of the reference samples is tested by comparing the differences between several groupings of reference samples into larger geographically arranged groups. This will test the differences between "ethnic" groups in several arrangements; firstly between the species included in the dataset, secondly between human groups based on their continent of origin, and thirdly between human groups separated on finer geographical criteria. The third section also includes the test on the influence of the average stature of the reference sample.

For overview purposes, all results of the statistical tests presented in this chapter have been summarized in table 5.12. The tests have already been presented in table 3.1, which is replicated here for convenience in table 5.1. The stars behind the p-values indicate a value between $p = .05$ and $p = .1$, whereas a value below $p = .05$ is indicated with two stars. In the test, only $p = 0.05$ is indicated, by one asterix. The standard deviation is abbreviated as SD or indicated by ± after an estimate.

Table 5.1: The tests executed on the reference sample hypotheses

Tests	**Tested by**	**Groups**	**Selection**	**Fossil**
1.1	T-test	females and males	groups with female and male equations	all
2.1	ANOVA	OLS, RMA, MA	groups with OLS, RMA and MA equations	OH 28
2.2	ANOVA	OLS, RMA, MA	idem, excluding logged equations	OH 28
3.1.1	T-test	nonhuman apes, humans	wide	OH 28
3.1.2	T-test	nonhuman apes, humans	raw-only	OH 28
3.1.3	T-test	nonhuman apes, humans	strict	all
3.2.1	ANOVA	Europeans, Africans, Asians, Americans	strict	all
3.2.2	T-test	Europeans, Africans	wide	all
3.2.3	T-test	Europeans, Africans	strict	all
3.3.1	ANOVA	Af. Am., Cau. Am., Ind. Am., mixed Am., Europeans, Af. Pygmies	strict	all
3.3.2	ANOVA	Af. Am., Cau. Am., Ind. Am., Europeans, Af. Pygmies	strict	all
3.3.3	ANOVA	Af. Am., Cau. Am., Ind. Am., Europeans	strict	all
3.3.4	ANOVA	Af. Am., Cau. Am., Ind. Am., Europeans	wide	all
3.3.5	T-test	Af. and Cau. Am.	strict	all
3.3.6	T-test	Af. and Cau. Am.	wide	all
3.4.1	ANOVA	size groups	wide	all

5.1 Data

The dataset used here consist of the stature estimation equations and their results for a range of input values. The equations can be divided into several groups following their reference samples and mathematical transformation. There are 6 log-transformed equations in the sample, which amounts to 5.3% of the sample. The nonhuman group make up 12.4% of the total sample with 14 equations. The Pygmy group consists of 13 groups, 11.5% of the total. Ratio equations make up 5.3% of the sample with 6 equations. Half of these are human-based, the other half are the ratio equations for three nonhuman hominoid species.

The total dataset which is worked with consists of 110 formulae, though 118 equations are present in the database. The 8 equations which are excluded from the total dataset are equations which calculate skeletal stature instead of living stature. They are included in the database because these are the original equations as published (Dayal *et al.* 2008; Feldesman *et al.* 1989; Sciulli and Giesen 1993) which have been adjusted by me to give living stature results.

For certain of these tests, the total dataset of equations was subjected to a strict selection of the equations, in order to include only one representative equation per reference sample. For this "strict selection" the excluded equations were: all female-based equations, all RMA and MA based equations when OLS was available, and all combined-sex equations when a male only - based equation was available. This restrictive selection is made to exclude all equations possibly creating noise because they differed in other factors too (e.g. mathematical method), and to minimize the extra weighing of the reference samples on which many equations were based. On the opposite side, the "wide selection" includes all relevant equations without applying a selection apart from e.g. "human" when only human-based equations are to be tested.

The test results differ depending on which fossil is used as input. The choice has been made to use OH28 as standard input in all tests. This fossil has been chosen because of its femur length, 45.6 cm, which is close to the femur length mean in many modern human populations (see appendix). In addition, some tests have been executed on the smallest fossil in the sample (KNM-ER 1463, 31 cm), the largest fossil in the sample (Berg Aukas, 51.8 cm) and on D4167 (38.6 cm) because of its position in between the value of KNM-ER 1463 and OH 28. The entire range of the fossil femoral lengths collected in this thesis is thus represented by these four discrete input values.

5.2 Results

5.2.1 Female - male influence

It is known that males and females form two stature subgroups in any human population, with males being consistently taller than females. It is recommended to use the sex-specific population equation for any goal individual within a population(e.g. Pearson (1899)). It is therefore hypothesized that equations based on males and females give different results.

The null hypothesis is that the means of the male and female based equations are not significantly different. This is tested by an independent samples T-test on the means of male and female equation results. The equation selection used in this test is formed by the reference samples for which both a female and a male equation are available. The test was executed using the femur length input of OH 28.

The results of this test can be seen in table 5.2. The T-test gives a 2-tailed significance (for assumed equal variances) of $p = .262$ (t(36) = -1.139). This test on the smallest (31 cm) and largest (51.8 cm) femur in the sample revealed the same consistent difference between the mean result for males and females; the male mean is always around 2 cm larger than the female mean while the difference is not significant. The null hypothesis could not be rejected.

Table 5.2: Test 1.1: The combined results for female-only and male-only equations for OH 28 (45.6 cm), KNM-ER 1463 (31 cm) and Berg Aukas (51.8 cm)

Fossil	**Sex**	**N**	**Mean in cm**	**SD**
OH 28	Female	19	165.7	2.5
	Male	19	167.3	5.7
KNM-ER 1463	Female	19	130.6	
	Male	19	132.5	
Berg Aukas	Female	19	180.4	
	Male	19	182	

5.2.2 Mathematical method influence

There are several mathematical methods available for the calculation of a regression equation. The most commonly used is ordinary least squares (OLS), but major axis (MA) and reduced major axis (RMA) are also used. It is tested whether the choice of mathematical method used for calculating the estimation equation has influence on the results.

The null hypothesis is that the means of the results based on different mathematical methods are not significantly different. This is tested by a one-way ANOVA. Excluded from this analysis are all equations from publications that do not explicitly test differences in mathematical method. Included are data from Feldesman and Lundy (1988); Hens *et al.* (2000); Jungers (1988); Maijanen and Niskanen (2010). The fossil that is used as input is OH 28.

In the first test, the equations used are based on log-transformed and on raw data. The ANOVA results in insignificance at $p = .906$ ($F(2,31) = .099$), see table 5.3. The groups show low within-group coherence, as seen by SD's of 14.5, 25.1 and even 39.2 cm.

When the test is run on raw data only, the ANOVA still results in insignificance at $p = .475$ ($F(2,25) = .767$). The within-group coherence is somewhat better without the log-transformed equations, as can be seen in table 5.3 as well. The null hypothesis could thus not be rejected in either test.

Table 5.3: Tests 2.1, 2.2: The combined results for the three different mathematical techniques (OLS, RMA, MA) used to calculate the estimation regressions, for OH 28 (45.6 cm)

Data included	**Mathematical method**	**N**	**Mean in cm**	**SD**
Raw and log	OLS	13	175.3	14.5
	RMA	11	171.3	25.1
	MA	10	176.2	39.2
Raw only	OLS	11	172.6	11.6
	RMA	9	176.9	18.6
	MA	8	185.1	33.5

5.2.3 Reference sample influence

The main hypothesis is that reference samples influence the results of estimation equations. While there is no doubt that each population achieves highest accuracy in stature estimation with an equation specific to that population, there are questions about the mechanism behind this. It could be that a person is more accurately estimated with an equation based on his/her population because he/she is more closely related to these people and therefore more similar to them in body proportions. It could also be that these people are similar in body proportions because they grew up in the same environment or because they are of similar total stature.

Not all of these possible explanations are tested here. Four sets of hypotheses are tested. The first concerns the difference between results of equations based on humans and based on nonhuman apes. The second tests the differences between continent-based origin groups.

The third uses another classification scheme based on combined geographical and ethnic criteria. The fourth hypothesis tests the difference between equations based on total stature groups. The null hypothesis will be that different groups of reference samples do not have a significantly different mean.

Human and nonhuman reference samples

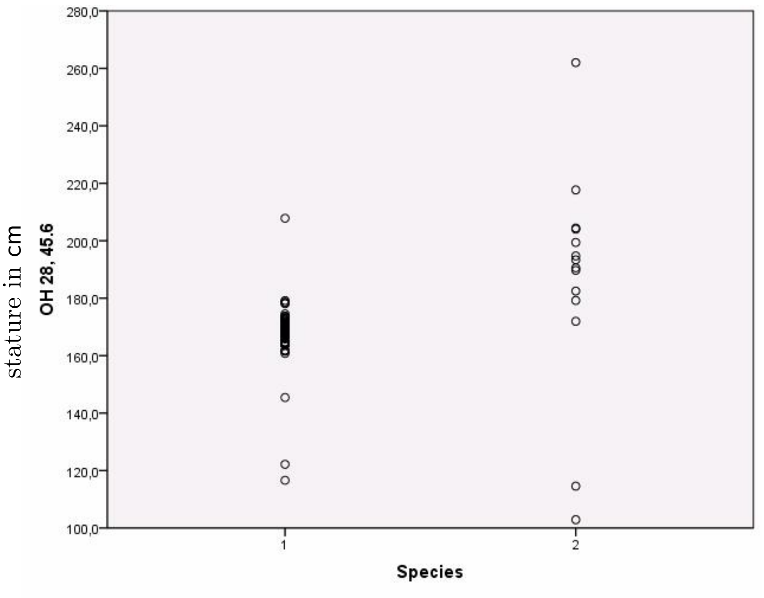
It is tested whether equations based on nonhuman reference samples give different results than equations based on human-only reference samples. Several independent samples T-tests are done on these two groups. The first test compares all 14 ape-based and 80 human-based equations. The second test compares raw data-only ape-based and raw data-only human-based equations. A third test is done on a strict selection of ape- and human-based equations. The first three tests are executed on the input variable of OH28. The last test is also executed on KNM-ER 1463, D4167, and Berg Aukas.

Table 5.4: Tests 3.1: The combined results for the equations based on humans and on nonhuman apes, for OH 28 (45.6 cm), KNM-ER 1463 (31 cm), D4167 (38.6 cm) and Berg Aukas (51.8 cm)

Fossil	**Sample**	**N**	**Mean in cm**	**SD**
OH 28	Ape log + raw	14	186.2	39.3
	Human log + raw	80	167.7	10
	Ape raw	11	201.6	22.7
	Human raw	77	166.9	8.9
	Ape strict	5	189.8	9
	Human strict	40	168.7	4.9
KNM-ER 1463	Ape strict	5	127.3	3.8
	Human strict	40	131.5	8.9
D4167	Ape strict	5	159.9	6.6
	Human strict	40	150.9	5.3
Berg Aukas	Ape strict	40	184.4	6.9
	Human strict	5	216.4	12.9

When the whole nonhuman sample (14 cases) is compared to the human sample (80 cases), the differences are insignificant when equal variances are not assumed: $T(13.3) = -3.664$, $p = .103$. When the raw-only nonhuman sample is compared to the raw-only human sample, 11 nonhuman equations remain. When equal variances are not assumed, Levene's test is significant, $T(10.4) = -5.003$, $p = .000$*.

Removing the log-transformed equations from the comparison also improves the within-group coherence, which is especially reduced in the ape sample. In this group, the SD for all equations is 39.3 cm but 22.7 cm for raw-only. Compared to the human SD (see table 5.4) it is still high, but unlike the human group the ape group is composed of more than one species, so a larger variation would be expected. The spread plots, figure 5.1 and figure 5.2, visualize this pattern. Figure 5.1 shows all equations spread over the two groups. Figure 5.2 shows the same, but without the log-transformed equations. From this it follows that removing the log-transformed equations reduces the amount of spread, especially within the ape group.



Legend:

1 Humans

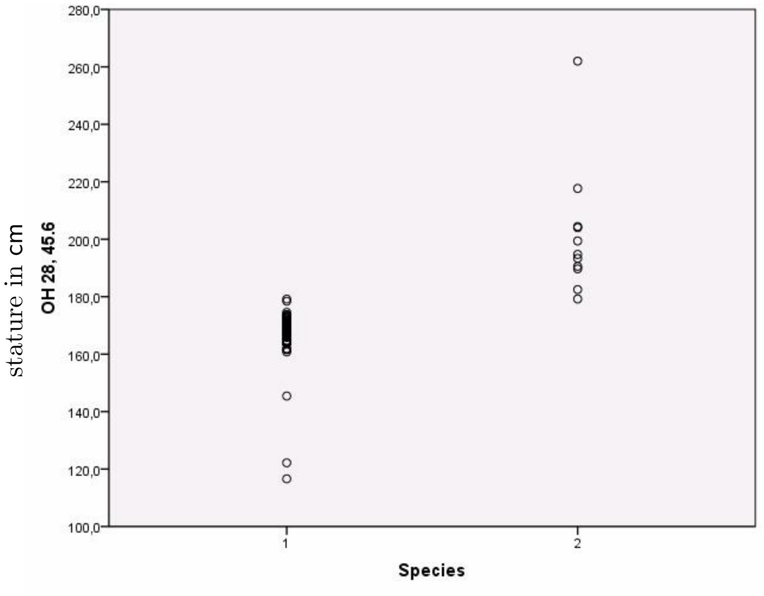
2 Nonhuman apes

Figure 5.1: Range plot per species for OH28

The restrictive selection yields 40 available equations for the human group (M= 168.7, SD 4.9) and 5 (M= 189.8, SD 9.9) for the nonhuman group, see table 5.4. Levene's test is significant, so unequal variances are assumed, resulting in $T(4.3) = -4.702$, $p = 0.08$*.

Comparing these two groups for KNM-ER 1464 results in insignificance $T(43) = 1.022$, $p = .313$, but in significance for D4167 ($T(43) = -3.44$, $p = .001$*) and for Berg Aukas ($T(43) = -8.78$, $p = .000$*). All of these are non-significant on Levene's test. The human - nonhuman ape groups thus give significantly different results for the largest fossils, but not for the smallest fossil.

Range plots of the results per species visualize why in the smallest fossil the groups are not significantly different. The range of results from ape and human equations are overlapping strongly in KNM-ER 1463, as can be seen in figure 5.3. On the contrary, Berg Aukas as the largest fossil in the dataset does not show overlap in results between human and ape-based equations, see figure 5.4. The spread in the human results is also somewhat larger in KNM-ER 1463, though not as large as implied in the graph: 45 cm in KNM-ER 1463, 40 cm in Berg Aukas.

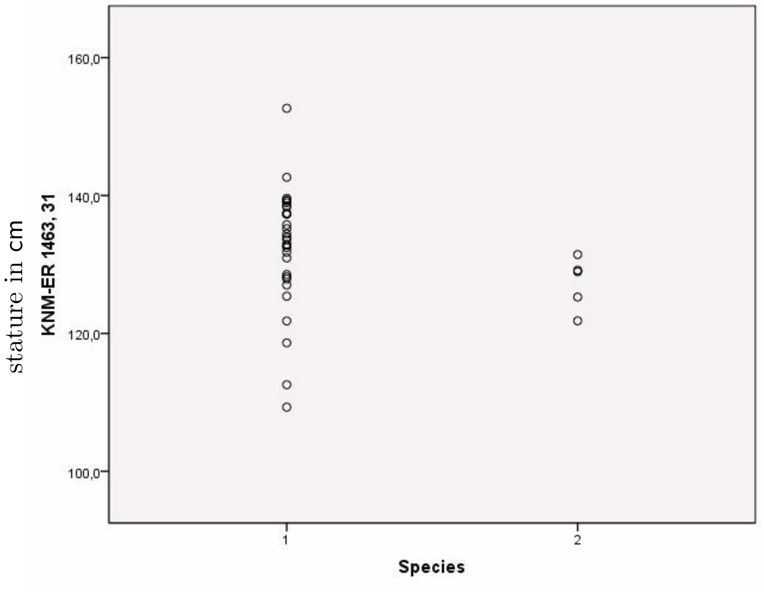


Legend:

1 Humans

2 Nonhuman apes

Figure 5.2: Range plot per species for OH 28 excluding the log-transformed equations



Legend:

1 Humans

2 Nonhuman apes

Figure 5.3: Range plot per species for KNM-ER 1463 using the strictly selected dataset

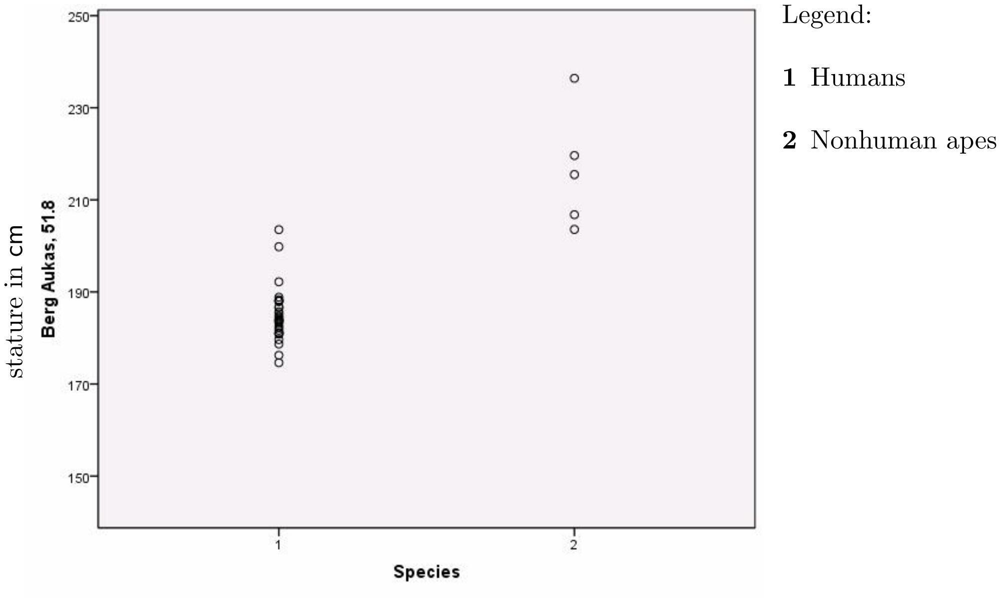


Figure 5.4: Range plot per species for Berg Aukas using the strictly selected dataset

The conclusion drawn from this is that ape-based equations overall give different results from human-based equations. H0 is rejected. This does not hold for the fossils in the smaller range. Stature estimations for fossils in the femur length range of apes are similar for ape-based and human-based equations.

Continent-based origin groups

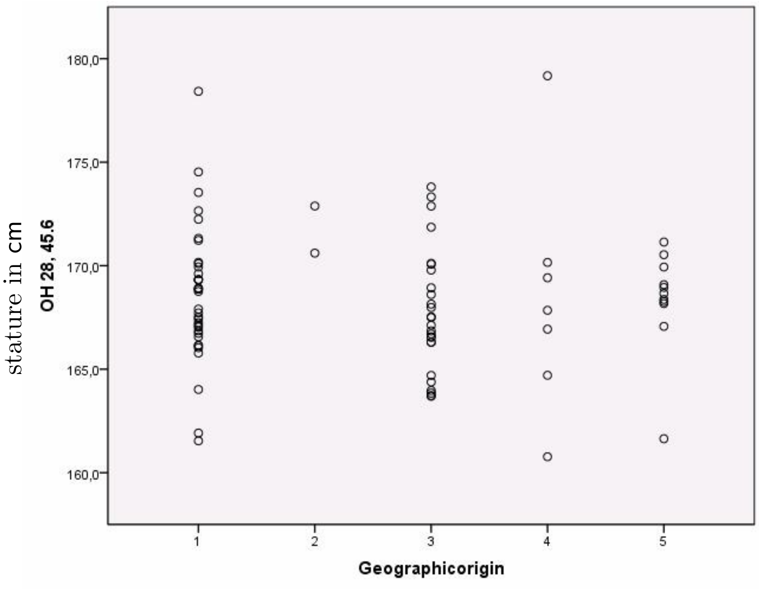
The equations were separated according to the continent where the samples originally came from. This resulted in four human groups on which ANOVA could be executed: European, Asian, African and Indigenous Americans. A fifth group, mixed populations, was left out of the ANOVA analysis.

When the strict selection is applied (one equation per reference sample only, N = 32) and ANOVA is executed on the four groups, this results in insignificance for all four fossils. OH 28 (Levene insignificant, $F(3,28) = 1.023$, $p = .397$), KNM-ER 1463 (Levene insignificant, $F(3,28) = 1.553$, $p = .223$), D4167 (Levene insignificant, $F(3,38) = 2.312$, $p = .098$) and Berg Aukas (Levene (only just) insignificant ($p = .064$), gives $F(3,28) = .297$, $p = .827$. A Welch test, executed because the Levene test was almost significant, resulted in insignificance as well ($p = .881$). Descriptives are presented in table 5.5. From these tests it is concluded that H0 cannot be rejected; means of groups are not so different that they could not come from one population.

The spread of equations per group can be seen in figure 5.5 (all equations) and figure 5.6

Table 5.5: Test 3.2.1: The combined results following the continent-of-origin classification (strict selection), for OH 28 (45.6 cm), KNM-ER 1463 (31 cm), D4167 (38.6 cm) and Berg Aukas (51.8 cm)

Fossil	Sample	N	Mean in cm	SD
OH 28	Europeans	17	169.6	3
	Asians	2	171.7	1.6
	Africans	9	167.8	2.7
	Americans	4	170.1	6.4
KNM-ER 1463	Europeans	17	135.1	6.3
	Asians	2	138.3	1.4
	Africans	9	129.7	9
	Americans	4	130.9	7.5
D4167	Europeans	17	153	4
	Asians	2	155.7	.2
	Africans	9	149.7	3.3
	Americans	4	151.2	3.2
Berg Aukas	Europeans	17	184.3	3.7
	Asians	2	186	2.9
	Africans	9	183.8	7.1
	Americans	4	186.9	11.1



Legend:

1 European origin

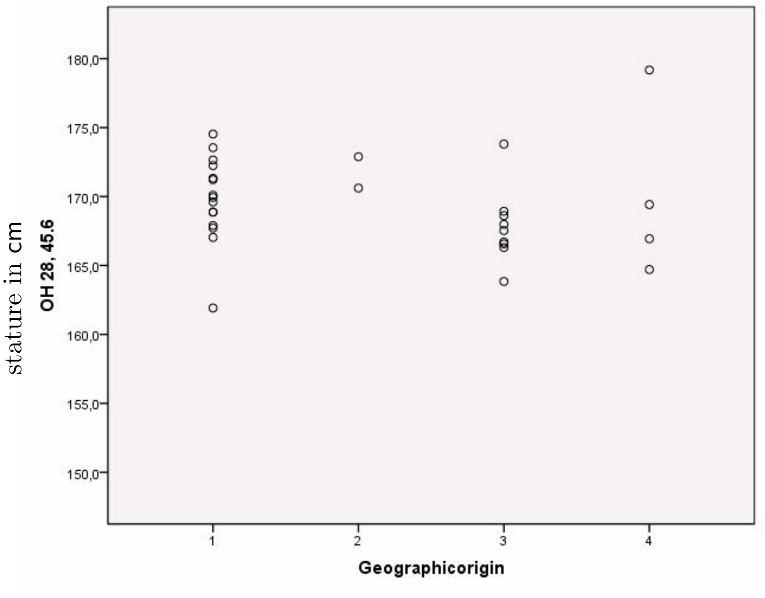
2 Asian origin

3 African origin

4 American origin

5 mixed populations

Figure 5.5: Range plot for continent groups using the wide selection



Legend:

1 European origin

2 Asian origin

3 African origin

4 American origin

Figure 5.6: Range plot for continent groups using the strictly selected dataset

(strict selection). Spread is between 160 and 180 cm for 45.6 cm femur length in all groups. The internal spread is largest within the indigenous American group, but otherwise seems to be relative to the group size. In figure 5.6 a strict selection of equations is applied. Though the number of equations is reduced from 88 equations (all human equations) to 32, the absolute spread is only slightly affected as it is still between 160 and 180 cm for all groups together. The European group for example still shows a high internal consistency, with almost all strictly selected equations clustering within 10 cm of each other. Looking at the means for the groups (table 5.5), they are within 10 cm and for most fossils even within 5 cm apart.

As becomes clear from table 5.5, only two of these continent groups have sufficient N, Europeans and Africans. An independent sample T-test on the wide selection of groups (N=65) results in $t(63) = .232$, $p = .817$ for OH 28, $t(45.5) = 3.53$, $p = .001^*$ for KNM-ER 1463 (Levene significant), $t(63) = 2.2$, $p = .031^*$ for D4167, and $t(32.4) = -1.099$, $p = .280$ for Berg Aukas (Levene significant). See table 5.6.

When the same test is applied to a strict sample (N=26), results are: $t(24) = 1.67$, $p = .107$ for OH 28, $t(13.2) = 2.06$, $p = .059$ for KNM-ER 1463 (Levene significant), $t(24) = 2.59$, $p = .016^*$ for D4167, and $t(24) = -.031$, $p = .976$ for Berg Aukas.

Some of the input values thus give significant differences between the results of these two groups, while the results from other input values do not differ significantly between the groups. The results of the groups tend to differ more in the lower extremes of the input range. The strict and wide selection of equations do give different results, of which the strict

Table 5.6: Tests 3.2.2, 3.2.3: The combined results of the European and African continent-of-origin classification for OH 28 (45.6 cm), KNM-ER 1463 (31 cm), D4167 (38.6 cm) and Berg Aukas (51.8 cm)

Fossil	Sample	N	Mean in cm	SD
OH 28	Europeans	35	168.6	3.3
	Africans	30	168.1	12.7
1463	Europeans	35	134.3	8
	Africans	30	124.3	13.7
D4167	Europeans	35	152.2	5
	Africans	30	145.4	11.5
Berg Aukas	Europeans	35	183.2	4.1
	Africans	30	186.4	15.5
OH 28	Europeans strict	16	169.7	3.1
	Africans strict	10	167.7	2.6
KNM-ER 1463	Europeans strict	16	135.3	6.4
	Africans strict	10	127.7	10.6
D4167	Europeans strict	16	153.2	4.1
	Africans strict	10	149	4
Berg Aukas	Europeans strict	16	184.3	3.7
	Africans strict	10	184.3	6.9

selection has preference due to theoretical reasons. The mixed picture that emerges from these results would indicate that at least for many input values the null hypothesis cannot be rejected, thereby increasing the likelihood that the continent of origin is not a biological relevant grouping factor for femur-based stature estimation equations.

Specific geographic origin

The mostly insignificant results between the continent-wide groups mentioned above might be due to the rough grouping criteria. In this way, populations that are not closely related are thrown together under the assumption that they are, and that close genetic alliance is the reason for any shared body proportions and for the likeness that their equations should have with each other. It is clear, however, that a continent scale can be too coarse for a genetic grouping. Therefore, it has been attempted to make biologically more relevant groups (see table in appendix). Some compromises had to be made due to scarceness of the data.

1 African Americans

2 Caucasian Americans

3 Asian origin: Americans (mixed Asian) and Asians

4 Indigenous Americans (both North and South Americans, including Inuit)

5 Mixed Americans (also mixed indigenous Americans - Caucasian)

6 Europeans

7 Africans

8 African Pygmies

9 Other humans

10 World wide humans

11 Nonhuman ape

12 Mixed human-nonhuman samples

The first test is executed on the groups African Americans, Caucasian Americans, Indigenous Americans, mixed Americans, Europeans and Pygmies. Levene's test of homogeneity of variances was significant for all fossils, so Welch tests instead of ANOVA were executed for all. This resulted for OH 28 in (5, 6.163, 2.247) $p = .173$, for KNM-ER 1463 (5, 6.541, 4.928) $p = .034$*, for D4167 (5, 8.951, 54.827) $p = .000$*, and for Berg Aukas (5, 6.138, 1.805) $p = .244$.

The second test is carried out on the groups African Americans, Caucasian Americans, Indigenous Americans, Europeans and Pygmies. The other groups are excluded because they have to few samples, or do not represent an actual population like the mixed groups.

This results in $F(4, 22) = 4.248$, $p = .014$* for KNM-ER 1463, $F(4, 22) = 3.674$, $p = .023$* for D4167 and a significant Levene test for both OH 28 and Berg Aukas, therefore warranting the use of the Welch test which scores *df* 4, 4.86, 2.13, $p = .157$ for OH 28 and *df* 4, 4.841, 2.125, $p = .219$ for Berg Aukas.

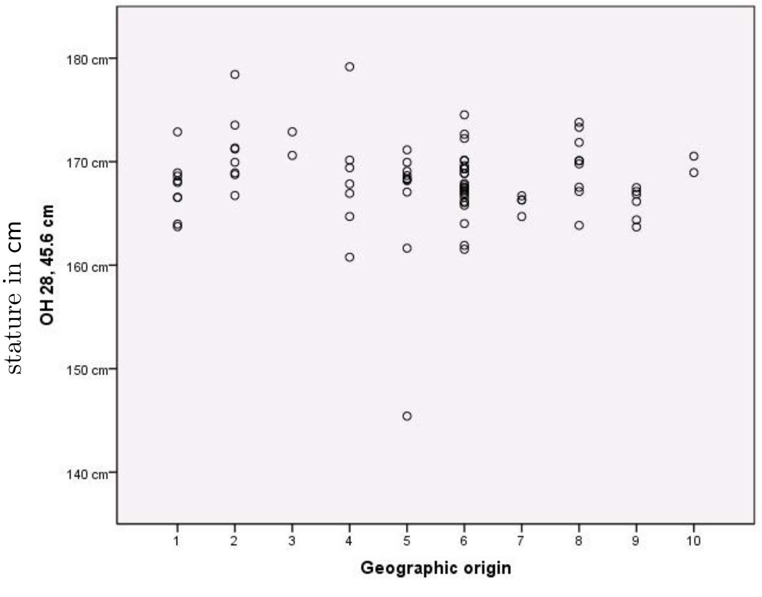
Looking at the mean results in table 5.7, the group responsible for this significant result appears to be the Pygmy group, which gives outlier results compared to the other groups for all fossils except OH 28. Next to representing an extreme in modern human stature, it is also a group consisting here of only 2 equations. The second test is executed with the same parameters, but excludes the Pygmy group from the comparison. The groups compared are African Americans, Caucasian Americans, Indigenous Americans, and Europeans.

Table 5.7: Tests 3.3.1, 3.3.2, 3.3.3: The combined results for the specific geographic classification (strict selection) for OH 28 (45.6 cm), KNM-ER 1463 (31 cm), D4167 (38.6 cm) and Berg Aukas (51.8 cm)

Fossil	**Sample**	**N**	**Mean in cm**	**SD**
OH 28	African Americans	4	168	1
	Caucasian Americans	4	171.5	1.5
	Indigenous Americans	4	170.1	6.4
	mixed Americans	5	164.6	10.8
	Europeans	9	169.7	2.4
	African Pygmies	2	170.5	4.7
KNM-ER 1463	African Americans	4	134.6	2.4
	Caucasian Americans	4	136.8	4.6
	Indigenous Americans	4	130.9	7.5
	mixed Americans	5	129	10.9
	Europeans	9	136.6	8.4
	African Pygmies	2	115.8	4.6
D4167	African Americans	4	152	.9
	Caucasian Americans	4	154.9	2.8
	Indigenous Americans	4	151.2	3.2
	mixed Americans	5	147.6	10.1
	Europeans	9	153.9	4.5
	African Pygmies	2	144.3	.2
Berg Aukas	African Americans	4	182.2	2.3
	Caucasian Americans	4	186.3	1.4
	Indigenous Americans	4	186.9	11.1
	mixed Americans	5	179.7	12.3
	Europeans	9	183.8	4.2
	African Pygmies	2	193.7	8.7

The results are insignificant when the values of KNM-ER 1463 are compared ($F(3,20) = .722$, $p = .552$), as it is for D4167 ($F(3,20) = .928$, $p = .449$, Berg Aukas (Welch: *df* 3, 7.513, 2.838, $p = .110$) and OH 28 ($F(3,20) = .777$, $p = .523$).

When Pygmies are excluded, significance levels drop to very insignificant levels. Apparently, in the inclusion of equations based on small-bodied individuals influences the ANOVA results to a large extent. This can be seen in range plot per specific geographic group, see figures 5.7, 5.8, 5.9, 5.10, and 5.11. The Pygmy-based group of equations tends to give more extreme results than the other human based equations, but this is only the case for the extremely sized (smallest, largest) fossils.



Legend:

1 African Americans

2 Caucasian Americans

3 Asian origin

4 Indigenous Americans

5 Mixed Americans

6 Europeans

7 Africans

8 African Pygmies

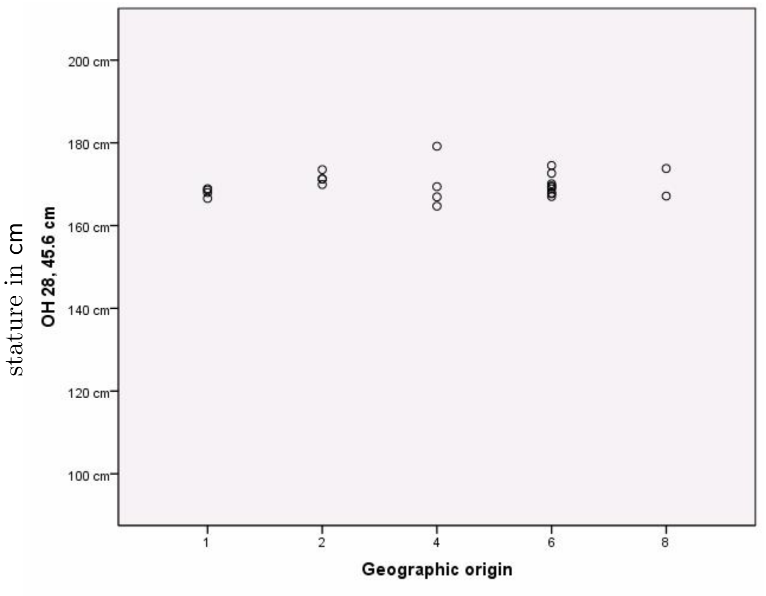
9 Other human groups

10 World wide

Figure 5.7: Range plot for geographic groups using the entire dataset for OH 28

To check if the strict selection of equations influences the results, the same ANOVA is executed on a less strict selection of equations ($N = 49$). The groups included are African Americans, Caucasian Americans, Indigenous Americans, and Europeans. Two of the tested fossils resulted in significant differences between the four groups, while the groups were not significantly different with the data of the other two fossils but did have lower p-values than in the strict-selection test. Significance was attained in OH 28 ($F(3,45) = 2.87$, $p = .047^*$) and D4163 ($F(3,45) = 3.51$, $p = .023^*$), but not in KNM-ER 1467 ($F(3,45) = 2.48$, $p = .073$) and Berg Aukas ($F(3,45) = .607$, $p = .614$). The descriptives are published in table 5.8.

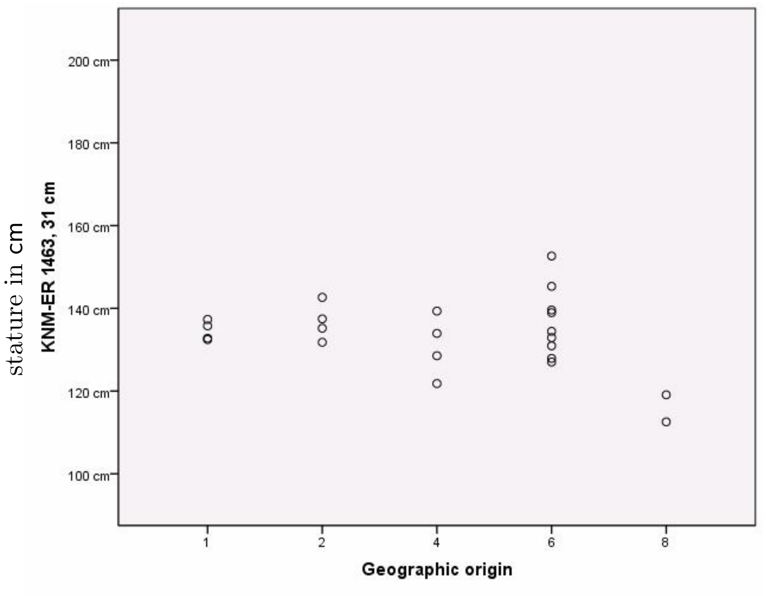
Surprisingly, the patterns in significant difference in this test differ from those in the tests on the strict dataset. The middle-sized fossils give significant results; these include the



Legend:

- **1** African Americans
- **2** Caucasian Americans
- **4** Indigenous Americans
- **6** Europeans
- **8** African Pygmies

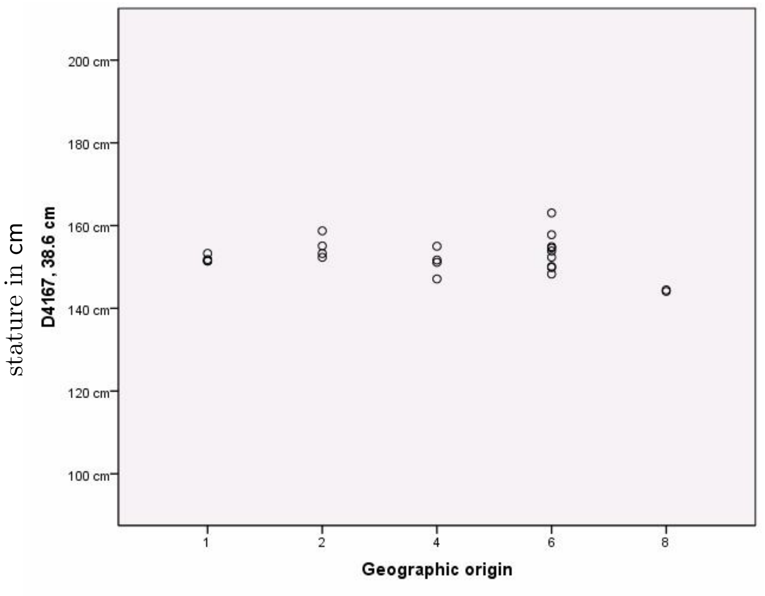
Figure 5.8: Range plot for geographic groups using the strictly selected dataset for OH 28



Legend:

- **1** African Americans
- **2** Caucasian Americans
- **4** Indigenous Americans
- **6** Europeans
- **8** African Pygmies

Figure 5.9: Range plot for geographic groups using the strictly selected dataset for KNM-ER 1463



Legend:

1 African Americans

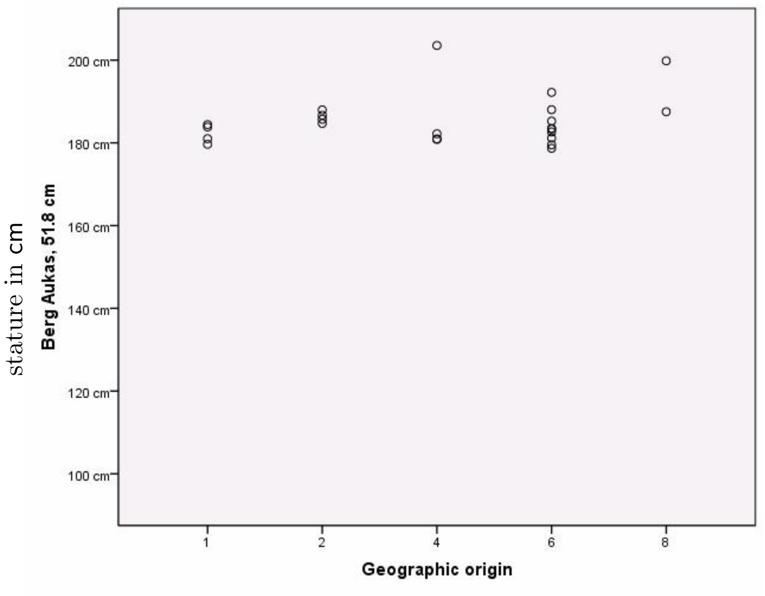
2 Caucasian Americans

4 Indigenous Americans

6 Europeans

8 African Pygmies

Figure 5.10: Range plot for geographic groups using the strictly selected dataset for D4167



Legend:

1 African Americans

2 Caucasian Americans

4 Indigenous Americans

6 Europeans

8 African Pygmies

Figure 5.11: Range plot for geographic groups using the strictly selected dataset for Berg Aukas

modern human sized femur OH 28. The extreme femora, both small and large, give insignificant differences between the population groups. Using a different selection of equations apparently does influence the results.

The internal consistency of the groups changes if the equation selection is less strict, as can be seen in table 5.8. For OH 28, the SD of Caucasian Americans goes up to 3.4 cm and that of Europeans to 3, while the SD of indigenous Americans reduces to 5.7 cm. The SD of African Americans, on the other hand, virtually remains the same at SD = 2.

Table 5.8: Test 3.3.4: The combined results for the specific geographic classification (wide selection) for OH 28 (45.6 cm), KNM-ER 1463 (31 cm), D4167 (38.6 cm) and Berg Aukas (51.8 cm)

Fossil	**Sample**	**N**	**Mean in cm**	**SD**
OH 28	African Americans	8	166.8	2
	Caucasian Americans	9	171.3	3.4
	Indigenous Americans	7	168.4	5.7
	Europeans	25	167.9	3
KNM-ER 1463	African Americans	8	130.3	8
	Caucasian Americans	9	140	11.2
	Indigenous Americans	7	131.2	6.3
	Europeans	25	133.4	7.4
D4167	African Americans	8	149.3	4.6
	Caucasian Americans	9	156.3	7
	Indigenous Americans	7	150.5	3.8
	Europeans	25	151.4	4.2
Berg Aukas	African Americans	8	182.3	2.8
	Caucasian Americans	9	184.6	2.4
	Indigenous Americans	7	184.4	8.9
	Europeans	25	182.6	4.5

An independent T-test (test 4) on the African and Caucasian Americans was executed. The data included are from the strict selection. The total N is 8, 4 in each group, see table 5.9. This test results in significance for the two largest fossils, OH 28 ($t(6) - 3.82$, $p = .009$*) and Berg Aukas ($t(6) - 3.02$, $p = .023$*), but not for the smallest fossils, D4167 ($t(6) = -1.94$, $p = .101$) and KNM-ER 1463 ($t(6) = -.85$, $p = .426$).

The last test of this hypothesis is an independent samples T-test like the previous test, but on a less strict selection in order to increase N (N=16). The resulting patterns in p-value is similar to the pattern seen in Test 4, though only one of the fossils tested shows significant differences between the groups. This is Berg Aukas ($t(14) = -3.44$, $p = .004$*).

Table 5.9: Test 3.3.5: The combined results from the T-test on African and Caucasian Americans (strict selection) for OH 28 (45.6 cm), KNM-ER 1463 (31 cm), D4167 (38.6 cm) and Berg Aukas (51.8 cm)

Fossil	**Sample**	**N**	**Mean in cm**	**SD**
OH 28	African Americans	4	168	1
	Caucasian Americans	4	171.5	1.5
KNM-ER 1463	African Americans	4	134.6	2.4
	Caucasian Americans	4	136.8	4.6
D4167	African Americans	4	152	.9
	Caucasian Americans	4	154.9	2.8
Berg Aukas	African Americans	4	182.2	2.3
	Caucasian Americans	4	186.3	1.4

OH 28 ($t(14) = -2.03$, $p = .062$), D4167 ($t(14) = -.902$, $p = .382$) and KNM-ER 1463 ($t(14) = -.462$, $p = .651$) are increasingly insignificant. See table 5.10

Table 5.10: Test 3.3.6: The combined results from the T-test on African and Caucasian Americans (wide selection) for OH 28 (45.6 cm), KNM-ER 1463 (31 cm), D4167 (38.6 cm) and Berg Aukas (51.8 cm)

Fossil	**Sample**	**N**	**Mean in cm**	**SD**
OH 28	African Americans	8	167.9	2.6
	Caucasian Americans	9	171.1	3.6
KNM-ER 1463	African Americans	8	135.8	8.4
	Caucasian Americans	9	138	10.1
D4167	African Americans	8	152.6	5.2
	Caucasian Americans	9	155.3	6.6
Berg Aukas	African Americans	8	181.5	2.3
	Caucasian Americans	9	185.2	1.8

Influence of mean stature

The fourth hypothesis is that the mean stature of a population influences the estimation equations formed on that population to such an extent that the estimation results will differ from other equations' results as a result of this. Results of equations grouped according to average stature of the reference sample are thus expected to give significantly different results.

Average stature groups were created by coding size groups with random borders. For the descriptives per size group see table 5.11.

group 1 < 150 cm on average

group 2 $150 - 159.9$ cm

group 3 $160 - 170$ cm

group 4 > 170 cm

The ANOVA is executed with no restrictions placed on the equations which are incorporated. This gives N=60 as for a large number of equations no stature data were available. The comparison of size groups' results gives a significant result for the smallest fossil ($F(3,56) = 5.241$, $p = .003$*), but not for the others. D4167 ($F(3,56) = 1.120$, $p = .349$), OH 28 ($f(3,56) = .492$, $p = .689$) and Berg Aukas ($F(3,56) = 2.52$, $p = .067$) are show insignificant differences between the size groups.

Table 5.11: Test 3.4.1: The combined results per stature group (wide selection) for OH 28 (45.6 cm), KNM-ER 1463 (31 cm), D4167 (38.6 cm) and Berg Aukas (51.8 cm)

Fossil	**Sample**	**N**	**Mean in cm**	**SD**
OH 28	< 150cm	11	170.7	21.1
	150 – 160cm	9	165.5	1.9
	160 – 170 cm	23	168.5	2.8
	> 170 cm	17	166.5	11.6
KNM-ER 1463	< 150 cm	11	116.6	14.3
	150 – 160 cm	9	136.6	9.5
	160 – 170 cm	23	132.7	7
	> 170 cm	17	130.7	14.8
D4167	< 150 cm	11	144.8	17.7
	150 – 160 cm	9	147.4	3.4
	160 – 170 cm	23	151.5	3.8
	> 170 cm	17	149.3	12.7
Berg Aukas	< 150 cm	11	193.6	24.1
	150 – 160 cm	9	181.5	5.1
	160 – 170 cm	23	183.6	3.6
	> 170 cm	17	181.8	11.4

5.3 Analysis of the statistical test results

5.3.1 General patterns

Presented in figure 5.13 are all results for all equations for a given input value range of 30 to 52 cm (standardized dataset). The pattern that is already visible in this figure, is even more clear when a larger input range is taken as in figure 5.14. The input values of that last figure are well beyond the natural femur length input range, but the figure is very instructive to study the behaviour of the equations and therefore included.

This "standardized dataset" is used aside of the "fossil dataset" for overview purposes; the input values used are placed at a regular interval instead of the random input values of the fossils. A larger input range can be seen in figure 5.14. These figures show per input value the spread of results by means of boxplots. The box represents the middle 50% of all data points, the whiskers cover the lowest and highest 25% (quartiles) of all data points. However, not all datapoints are covered by either box or whiskers. These are indicated by circles and stars, which are outliers (circles) and extreme outliers (stars). The circles represent results which are more than 1.5 box lengths away from the closest box border, the stars are more than 3 box lengths away form the same point. This is explained clearly in figure 5.12. The numbers attached to the outliers correspond to the numbers of the formulae in the SPSS dataset (see appendix).

From this follows that a smaller box means that the middle quartiles' datapoints are more concentrated than a larger box and similarly for the length of the whiskers for the other quartiles' datapoints. The pattern visible in both figure 5.13 and especially figure 5.14 shows that the concentration of results is highest around 44 - 46 cm femur length, while the results are more evenly spread for input values away from this point. This means that the equations converge around 45 cm, which is the smallest box. Why do most results cluster closely around 45 cm, while they show a wider variance with other femur lengths? Does this say anything about the reliability of the results?

The hypothesized explanation for the clustered spread of results of the femora around 45 cm in relation to the larger overall variance of results of smaller and larger femora is the usage of reference samples with femora that average around 45 cm. For example, the average femur length and stature for modern humans in the large mixed American sample in Hens *et al.* (2000) is 46.6(±2.8) cm and 172.5(±8.5) cm (modern Americans are relatively tall in comparison with some of the other groups). Equations formed on this information perform well around the average and within the boundaries of the reference sample, so most equations based on modern humans give similar answers around the average modern human femur length. Outside this average, the different specifics of each equation (intercept and slope), which are fine-tuned only to give matching results in the sample range, give

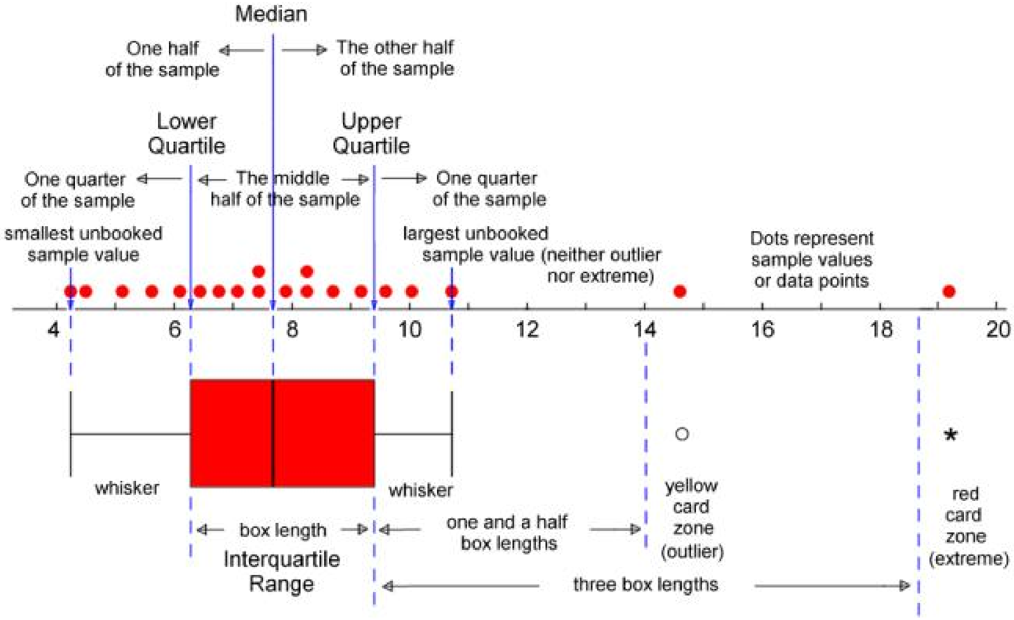


Figure 5.12: Explanation of boxplot meaning, courtesy of Murdoch University (Murdoch University 2009)

increasingly differing results. This is the reason that extrapolation outside the value range of the reference sample is dangerous. This interpretation is strengthened when the boxplots are split into separate figures for females (figure 5.16) and males (figure 5.15). The females seem to converge on a lower femur length (42 cm) than the males (46 cm).

The sample of equations is dominated by modern humans of the 20th century who were analysed by OLS linear regression. More than 30% of this sample is based on European or otherwise Caucasian reference samples. Other “western” groups (African American, mixed Americans) account for 17% of the sample. Specifically small-bodied humans (both Pygmies and South African small-bodied individuals) are involved in 20% of cases, but a large percentage of this group could not be used in many of the analyses due to being combined with hominoids (20%), due to being log-transformed (13%) and due to repeated usage of the same reference sample. Hominoids are used in 12% of the equations. It is probably the “western” half of the data that is creating the visible pattern, forming the “box” in the boxplots; they represent a quarter of the outliers (6 out of 24 outliers). All other outliers are based on Pygmies (5 outliers), hominoids (12 outliers) or both (1 outlier).

The outlier position of a disproportionally large group of equations based on hominoids and Pygmies is not an unexpected thing. It indicates that their results are more different from the bulk of equations based on standard modern humans, than that these standard human-equations are to each other. As hominoids and Pygmies are included in stature

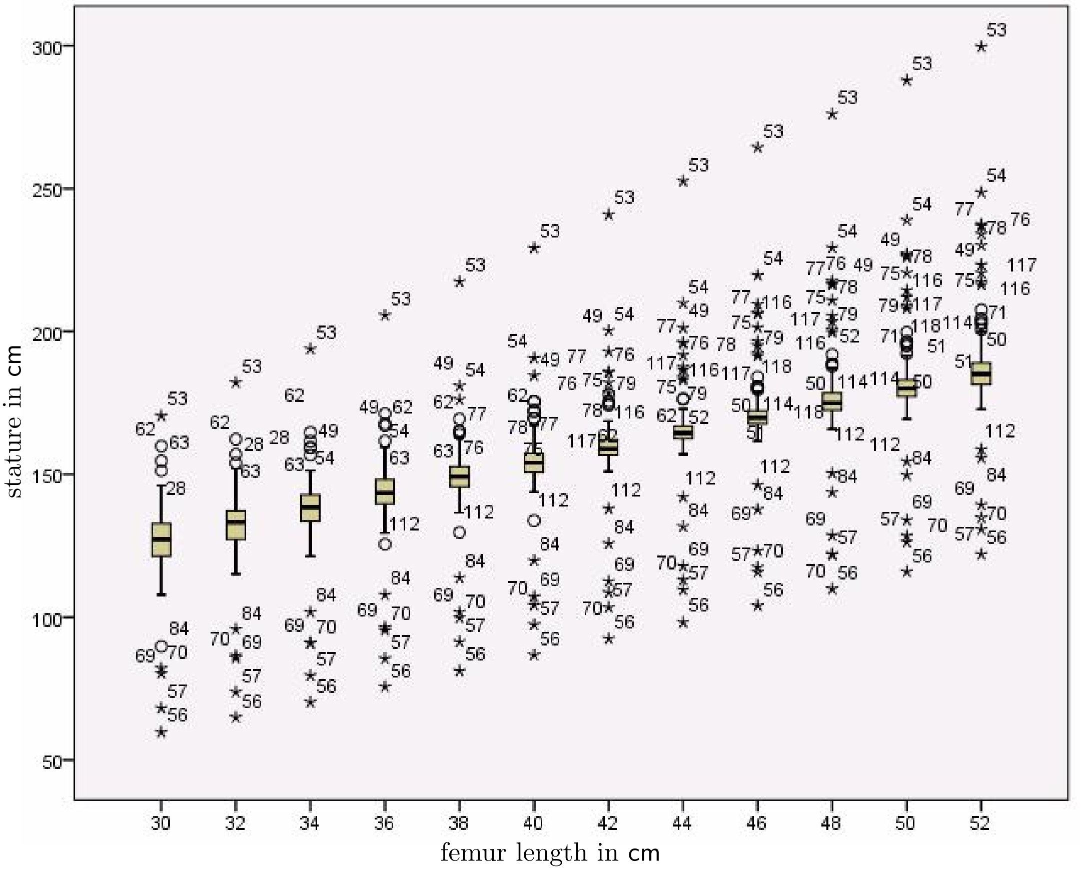


Figure 5.13: All equations for femur input range 30 - 52 cm

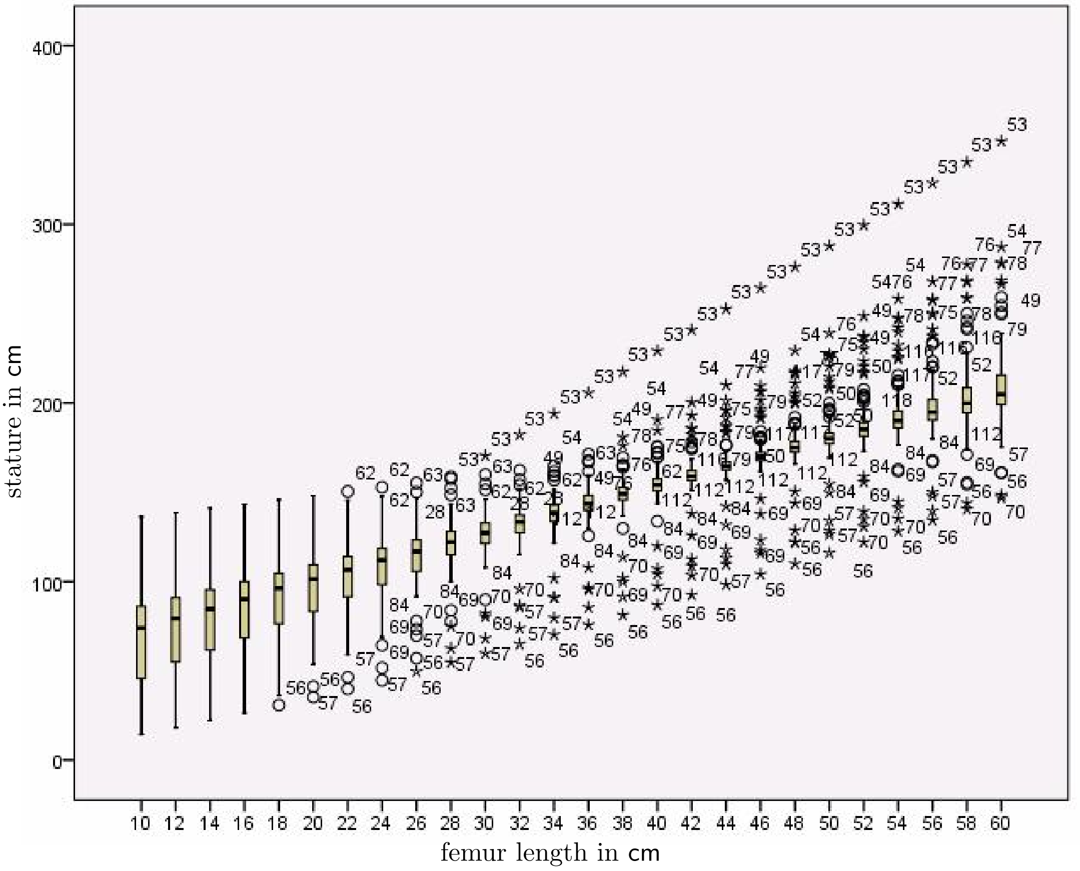


Figure 5.14: All equations for femur input range 10 - 60 cm

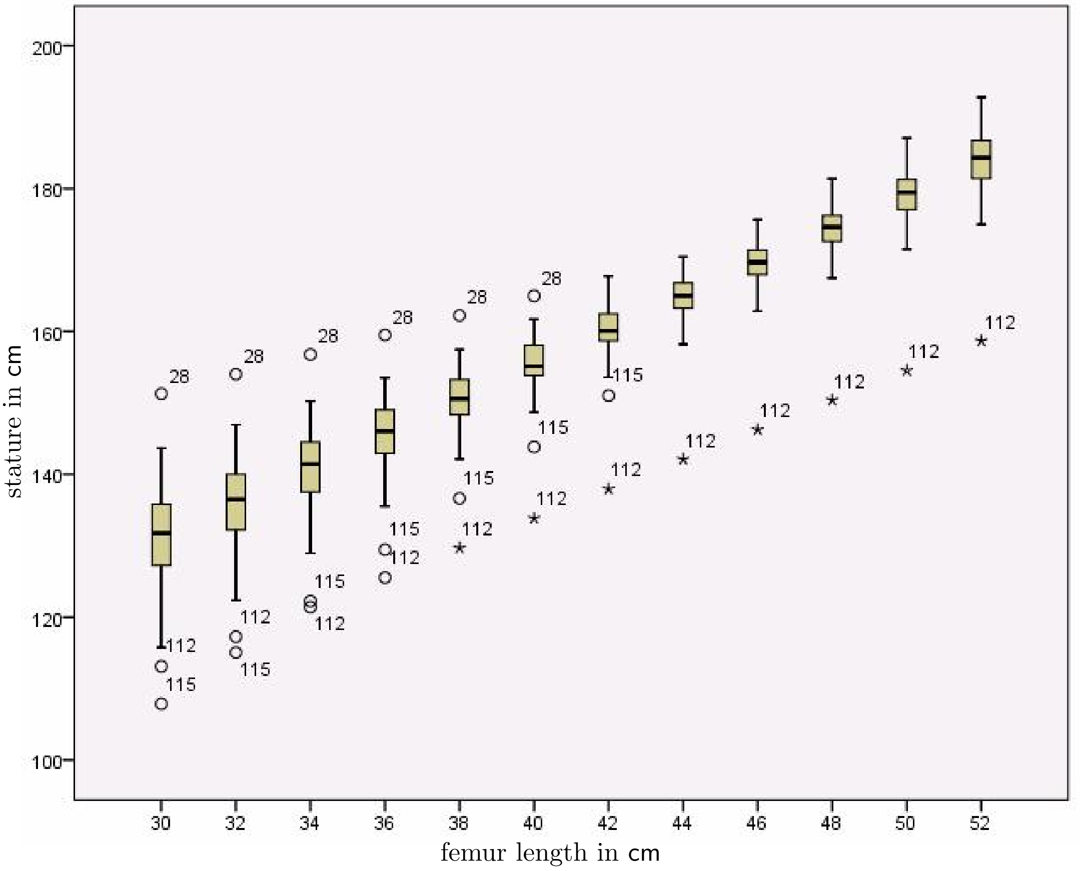


Figure 5.15: All male equations

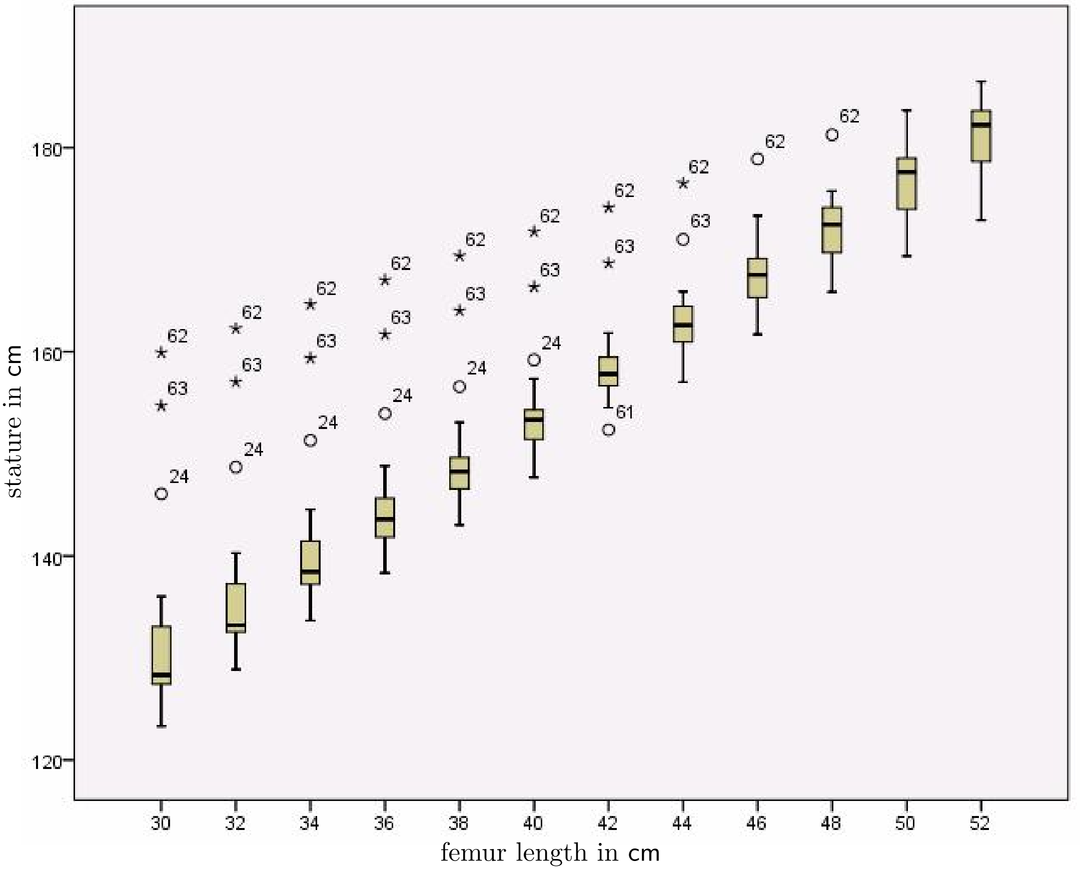


Figure 5.16: All female equations

estimation because their bodies differ from standard modern humans, this is not a surprise. Seen in this light, it is simply the selection of the equations which produces this picture. The distribution of outliers and clusters has therefore no direct bearing on the reliability of any of the results for fossils in the graphs. This centered distribution does not mean that the average of the centred equations in larger femora are more reliable estimates for the stature of hominins than the ones which are outside that focus; it means that modern humans with that size femur are likely to be of that stature.

A similar explanation as that given for the outliers in the general analysis above could be valid for the amount of non-linear equations presented as outlier. Log-transformed equations are only a small part of the sample (6 of the 110 equations), yet they are a disproportionally large group within the outliers. The high incidence of log-transformed regression in the outliers (5) indicates that the results of these different methods are not as similar to the "standard" linear non-logged regression methods results as the results of these standard methods are to each other, which would indicate that log-transformation is in fact a factor of influence in stature estimation. However, all logged equations are based on bonobos (3) and on Pygmies (3). These groups are expected and over represented amongst outliers, and could be the main reason for the outlier position of some of the log-equations.

Table 5.12: The statistical test results. * = between $p = .05$ and $p = .1$, ** = below $p = .05$, NS = not significant, *empty* = not tested

Test	**Tested by**	**Groups**	**Selection**	**OH 28**	**KNM-ER 1463**	**D4167**	**Berg Aukas**
1.1	T-test	females and males	female or male-based	NS	NS	NS	NS
2.1	ANOVA	OLS, RMA, MA	OLS, RMA and MA-based	.906			
2.2	ANOVA	OLS, RMA, MA	idem, excluding log-transformed	.475			
3.1.1	T-test	nonhuman apes, humans	wide	.103			
3.1.2	T-test	nonhuman apes, humans	raw-only	.000**			
3.1.3	T-test	nonhuman apes, humans	strict, raw-only	.08*	.313	.001**	.000**
3.2.1	ANOVA	4 continent groups	strict	.397	.323	.098*	.827
3.2.2	T-test	Europeans, Africans	wide	.817	.001**	.031**	.280
3.2.3	T-test	Europeans, Africans	strict	.107	.059*	.016**	.976
3.3.1	ANOVA	groups 1, 2, 4, 5, 6, 8	strict	.173	.034**	.000**	.244
3.3.2	ANOVA	groups 1, 2, 4, 6, 8	strict	.157	.014**	.023**	.219
3.3.3	ANOVA	groups 1, 2, 4, 6	strict	.110	.552	.449	.523
3.3.4	ANOVA	groups 1, 2, 4, 6	wide	.047**	.073*	.023**	.614
3.3.5	T-test	groups 1, 2	strict	.009**	.426	.101	.023**
3.3.6	T-test	groups 1, 2	wide	.062*	.651	.382	.004**
3.4.1	ANOVA	size groups	wide	.684	.003**	.344	.067*

5.3.2 Female - male influence

From the statistical analysis as presented in paragraph 4.2.2 it follows that there is no significant statistical difference between the results of female and male-based equations. It is clear however, that there is a biological pattern, visible in the consequent difference between female and male means of 2 cm. The differences that this pattern instigate are apparently overcome by the larger differences in total result range acquired by the compilation of 19 reference samples from around the world.

5.3.3 Mathematical method influence

The ANOVA tests on groups of equations separated based on mathematical method used were all insignificant. Several selections of data were tested, including and excluding the log-transformed equations.

When the logged equations are incorporated with the raw data-based equations, the averages of the three mathematical methods are within 1.5 cm of each other, reflected in the very high insignificance of the ANOVA result. The internal group coherence is not very good, as reflected in the SDs with somewhat biologically irrelevant levels of 14.5, 25.1, and 39.2 cm. In more restricted groups, e.g. raw data only groups, the averages are further apart (table 5.3) but the SDs remain quite high (*ibidem*) as does the insignificance of the ANOVA and/or, when appropriate, Welch results. Compared to the within-group coherence of the male/female-based equations in the last section, with SD's of 2.5 and 5.7 these mathematical method SD's are high.

This division does not give results which indicate that choice of mathematical method actually does have a large influence on the result. Other factors are more important. However, the data used here are first of all very limited with low N per group. Secondly, the groups included in the comparison have dissimilar reference samples, thereby complicating the comparison. These are modern humans (19) of which Pygmies (6), combined Pygmy-hominoid (3), and exclusively hominoid (6).

A visual representation of these equations can help to understand the pattern at hand. Figure 5.17 shows the trajectories of OLS, RMA and MA equations of three reference groups. This figure shows the human (Americans), Pygmy and hominoid equations as published by Hens *et al.* (2000). Three distinct groups are visible; all hominoids, Pygmies and Americans cluster together in groups, with the Pygmies and Americans closer to each other than to the hominoids. There is no grouping visible that has the mathematical method as factor. Instead, there is a clear divide between species. Within the groups, the same sort of clustering as in figure 5.13 can be seen. The humans and Pygmies equations cluster around 42 cm input, hominoids at around 32 cm input.

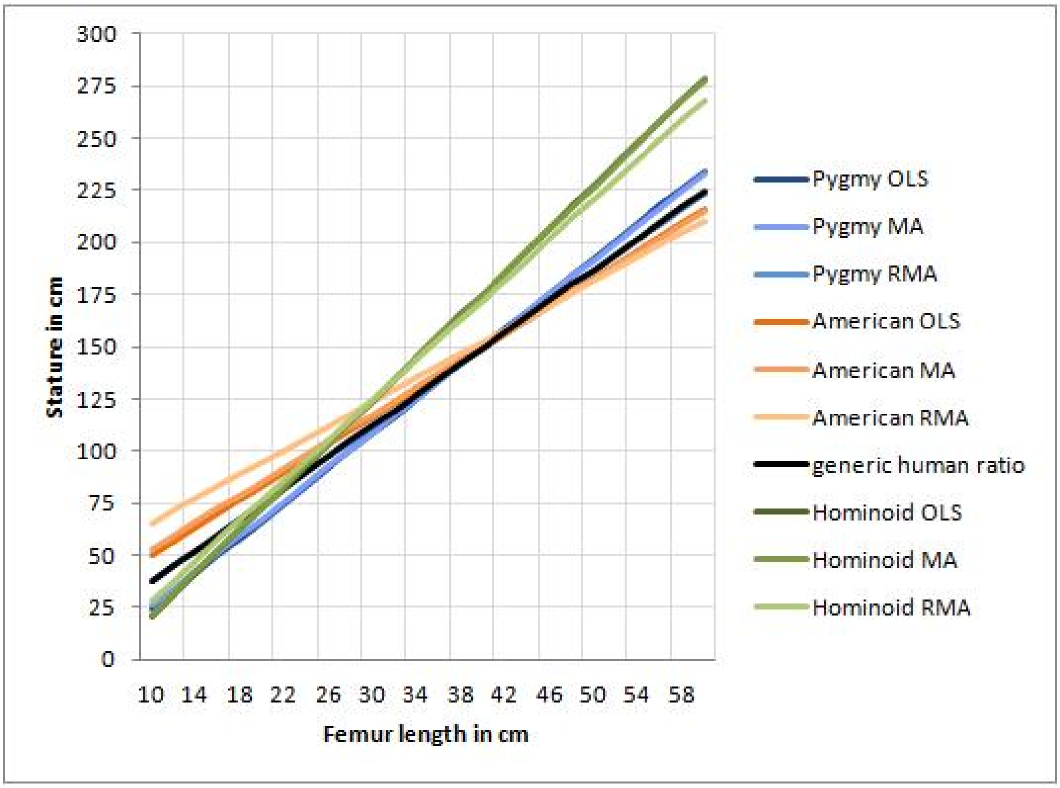


Figure 5.17: Mathematical method equations on Pygmy, American and Hominoid reference samples from Hens *et al.* (2000)

Ratio equations lack the intercept and are therefore usually steeper than linear regression equations, as is illustrated for example in figure 5.18, which shows several African and Caucasian American equation trajectories and two ratio equations. The result is that it gives lower results in the small input range, and larger results in the higher side of the input range.

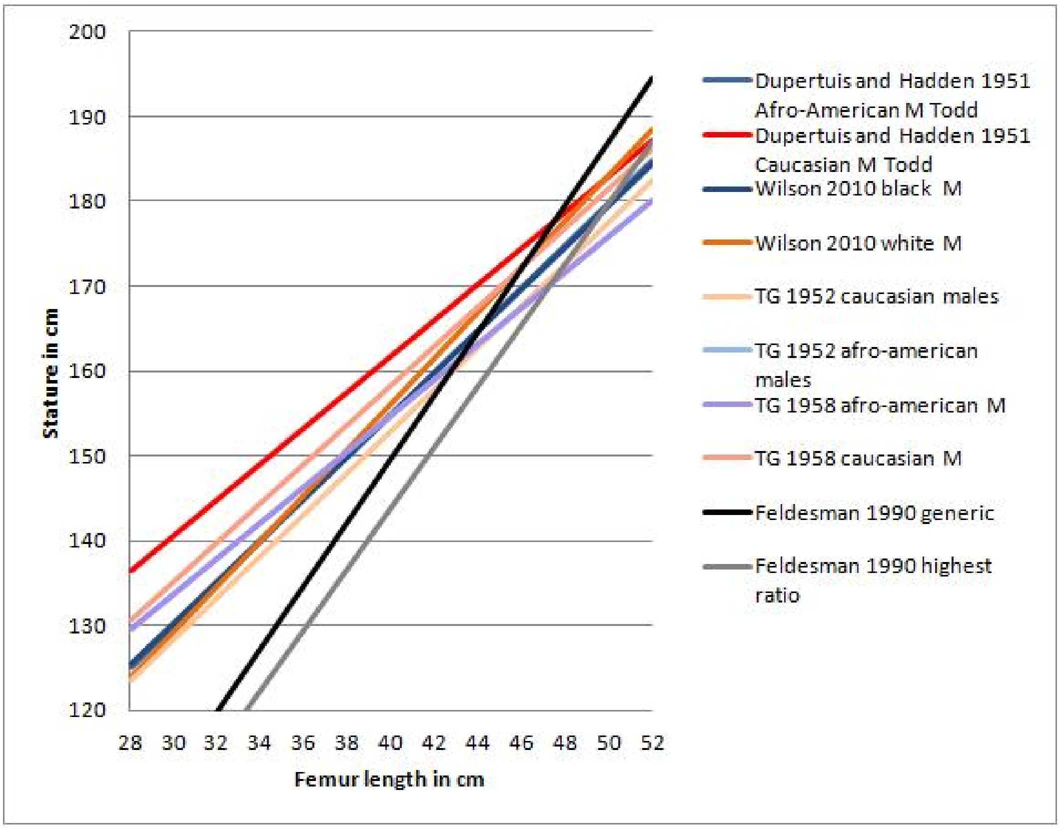


Figure 5.18: Several African and Caucasian American samples

5.3.4 Reference sample influence

Several hypotheses were formed to test aspects of the influence of reference samples on the estimation equation's result. The first hypothesis tested is: the results from equations based on different species will differ from each other. Four tests were executed to test this hypothesis, of which tests 3.1.2 and 3.1.3 result in significance for one or more fossils. The tests compared raw-only data in both the wide and strict selected datasets. Humans and apes are significantly different for OH 28, D4167, Berg Aukas, but not for ER 1464. Test 3.1.1 included also the log-transformed equations and resulted in insignificance.

The conclusion drawn from this is that ape-based equations overall give different results from human-based equations and that H0 can be rejected. The smaller fossils in the dataset cannot be included in this conclusion. The ANOVA does not give significant differences between human- and ape-based results for the smallest fossil in the range. An indication for the explanation of this deviation can be seen in figure 5.3 and figure 5.4 in combination with figure 5.13. As most human-based equations are made on a similar-sized reference sample with an average femur length of 45 cm and a minimum femur length of above 34 cm, a femur length of 30 cm is out of the reference sample range. Estimating body size for a femur length of 30 cm will therefore necessarily need extrapolation when these equations are used. As can clearly be seen in figure 5.14, the rule holds that the farther away from the mean the input value is, the wider apart are the results from the equations. This causes a wide spread of results at 30 cm input. The hominoid reference samples on the other hand do not need to be extrapolated, as the average femur length of the species used are 29.2, 29.4, and 34.6 cm. Their results therefore overlap with the wide-spread human results.

The comparison that included results of log-transformed ape-based equations give results that are not significantly different from the results of human-based equations. A look on figure 5.1 and figure 5.2, that shows the spread of the groups used in Test 3.1.1, explains why these groups are not different from each other. The spread of the results of the human and ape group overlap when log-transformed equations are included, but this picture changes as the log equations are removed (figure 5.2).

It appears that log-transformed equations based on hominoids give results more similar to human-based equations than to linear equations based on hominoids, in this case lower results. Though log-transformed equations are disproportionally present amongst (extreme) outliers, this does not have to be a reason to discard them completely; perhaps log-transformed equations based on hominoids are more suitable for the estimation of human-like subjects than linear equations based on hominoids.

The conclusion drawn is that ape-based equations give different results from human-based equations and H0 can be rejected. The interpretation from this is that grouping on species is a valid distinction. However, the small sample size in combination with other factors influencing the result (log vs raw data) makes the sample mean and T-test result swing easily. The validity of ignoring log-transformed equations is questionable, but they are left out of the following analysis to minimize noise. In the following tests, nonhuman apes were left out of the analysis in order to be able to concentrate on the differences within the human species.

The second hypothesis that is tested is that of the continent-of-origin of populations. A strict selection is applied to the equations, divided into four origin groups. This resulted in insignificance for all tested fossils, except a $p = .1$ significance for D4167. As only two of the groups had sufficient N, T-tests were executed on the European and African origin groups. This resulted in significant results for D4167 ($p = .05$) and KNM-ER 1463 ($p = .01$).

The internal consistency of the groups is quite high. The largest internal spread is within

Indigenous Americans, an observation that would fit in the light of the climatic adaptation hypothesis, as this is a large continent stretching North-South, covering all types of climate. The group includes Patagonian Indians but also Inuit.

It is not clear-cut which conclusion should be drawn from these statistical analyses as there seems to be some pattern present, distinguishing the results for the different groups in some fossils but not in others. This pattern is also visible in the results from the tests of the third hypothesis, that which tests the specific geographic groups. These results will now be discussed, after which it shall be attempted to draw a general conclusion.

The third test is executed on a finer-tuned grouping variable on the origin of the populations. A comparison is made on all groups with an $N > 1$. The internal consistency of the groups shows an expected pattern. The mixed American group has the highest SD, this is expected as it is not a uniform population but consists of a heterogeneous group of people from different populations. High internal consistency is seen in the African American group, the European group and the Caucasian American group. This high internal consistency was especially expected for the European group, which is a geographically restricted group. It is not due to over representation of one particular reference sample. The SD for the Pygmy and indigenous American groups are in between the low SD group and the mixed American group. This might be an effect of large geographical spread (indigenous Americans) or the low N (2, in the case of the Pygmies).

The initial ANOVA analysis resulted in insignificance for OH 28. The same analysis on the smallest fossil results in a significant difference between the groups. Excluding the mixed group in the ANOVA test (which is the theoretical right thing to do as they are not a real biological group) increases the significance-level for all tested fossils except OH 28. When the selection of groups is further restricted by excluding the pygmy group, significance levels drop to very insignificant levels. Apparently, the inclusion of equations based on small-bodied individuals influences the ANOVA results to a large extent. This can be visualized in range plots per geographic group, see figures 5.8, 5.9, 5.10, and 5.11. The Pygmy-based group of equations tend to give more extreme results than the other human based equations, except for OH 28.

A final T-test on the groups African Americans and Caucasian Americans resulted in an outcome with the groups significantly different from each other for the largest fossils but not for the smallest fossils. Figure 7.1 shows the trajectories of some of the equations of these groups. While there seems to be a pattern in which African American equations give lower results overall and especially for input above 42 cm, all equations tend to be clustered within 10 cm of each other.

Concluding from these tests it can be remarked that there appears to be some sort of pattern present, both in the continent-of-origin grouping as in the specific geographic grouping. Mostly, the smaller fossils give significantly different results while the large fossils only do so sometimes. OH 28 shows a much lower sensitivity to differences between the groups tested than the other fossils. It is proposed that this is due to it modern human-like size. As was concluded from figure 5.14 most of the results are concentrated around the mean of human femur lengths. OH 28 has a length of 45.8 cm, which is on this human femur mean. Therefore, on precisely this input value point lies a very small result range. This high concentration might prevent the groups of fossil to be distinguished from each other.

Both the continent grouping as the specific geographic grouping result for some of the fossils in significance. As this method of different groupings has been chosen as a test to see which grouping is relevant to stature estimation results, it is unexpected that two competing groupings both show significant results. However, the significant difference for both the continent-of-origin groups of European and African descent and the specific geographic groups of African and Caucasian Americans is perhaps unsurprising. The European descent group is partially made up of the entire Caucasian American group, similarly is the African American group part of the African descent group. Surprising then is the difference in the fossils where they give significance for; the continent groups for the smallest fossils and the specific geographic groups for the largest. I have not been able to find a good explanation for this difference.

The fact that both groupings show significant results may therefore be a consequence of an overlap between classification systems. After all, the specific geography grouping is a more detailed version of the continent-of-origin grouping. The more detailed version is more successful, as the ANOVA/Welch comparisons between 6 and 5 groups are significant while the ANOVA executed on the continent groups is hardly significant at all. The conclusion from this all could be that these specific geography reference sample based groups do give different results from each other, and H0 can be rejected when the groups compared are the Pygmy, Caucasian American, African American, Indigenous American and European groups.

A final interesting pattern to discuss is the difference between the ANOVA tests on specific geographic groups that includes and that excludes the Pygmy group. When the Pygmy, Caucasian American, African American, Indigenous American and European groups are compared with each other, this gives a significant result for the two smallest fossils. When the same groups without the Pygmy group are subsequently tested, this gives no significant results. Theoretically, there are good grounds to remove the Pygmies from the comparison: its N is only 2, and these two groups themselves are based on only few individuals. However, these groups are probably a good indication of the average stature of Pygmies, who are, of course, known for and defined on this attribute. Their status apart from the other groups is therefore not unexpected. It is also reflected in the result averages as seen in table 5.7.

The conclusion following from these results is that it is mainly the Pygmy group that is different from all other groups and that is responsible for the significant differences observed between the groups. All of the larger and more encapsulating groups that remain, Caucasian Americans, African Americans, Indigenous Americans and Europeans, do not differ from each other in a significant way. This could be due to the similarities in size between the humans in these groups, compared to the Pygmy individuals. It could also be due to the similarity in terms of genetic relatedness between the European and Caucasian American group. They have been grouped together in continent-of-origin group. Although it is known (Olivier 1963) that Europeans are overestimated when the Caucasian American formulae of Trotter and Gleser (1952) are used on them, they may still be not significantly different. Another test that might point in that direction is the T-test that compared two of the groups, African and Caucasian Americans, which are recognised as significantly different from each other for two of the tested fossils. No further tests were executed, so a final answer cannot be provided and the conclusion can go no further than recognizing the Pygmies as group standing apart from the other groups.

Size group influence

One thing that becomes clear is that the smallest stature reference samples do not necessarily give the smallest stature estimates. In fact, in the larger fossils it is the smallest average stature group that gives the highest estimates. Furthermore, a similar pattern as seen in other tests can be observed. The differences between the groups are smallest on the human-sized fossil (5 cm difference), and increase towards the femur input extremes. The largest differences between the groups are obtained in KNM-ER 1463, where it is as much as 20 cm.

5.3.5 Evaluation of the dataset selection

A choice has been made to limited the total dataset to a restricted dataset as the data for most tests. This was done to prohibit the over representation of certain reference samples and also to limit the possible influence of other factors. To evaluate if this made any difference, several tests were also executed on the wide (non-restricted) dataset. Three of those are used here for evaluation.

Two of the tests gave the same significance signal for the strict and wide dataset. This is when the Europe and Africa continent-of-origin groups were compared with each other, and when the African and Caucasian American geographic groups were compared. In the first instance, both the wide and strict test gave significant results for KNM-ER 1463 and D4167, while in the second case OH 28 and Berg Aukas were significant on both the wide and strict set. A reason for these similar results between the wide and strict tests could be a large overlap between the datasets, i.e., there are few equations that belong to the tested groups in the wide dataset that have been excluded from the strict dataset.

However, when four of the specific geographic groups were compared with ANOVA (test 3.3.3 and 3.3.4), the strict and wide selection gave very different results. The strictly selected dataset gave no significant results, while the wide dataset gave significant results for OH 28, KNM-ER 1463 and D4167. As this is a very relevant difference, I consider it an indication that making a selection of the dataset to purify it by minimizing noise and double-counting is useful in getting more meaningful results. The significant results of the wide dataset may be false creations of larger within-group coherence manufactured by over representation of certain samples as N is not actually increased when a sample is presented several times. The low N in the strict dataset however, may inhibit the recognition of actual groups. This cannot be resolved without the addition of more reference samples which fit with one of the groups tested.

5.3.6 Conclusion

What do these tests say on the suitability of the estimation techniques for *Homo erectus*? The hypothesis that got the most support from the statistical tests was the specific geography classification. However, this was related to the presence of the Pygmy group in the analyses. In combination with the significant results obtained with the stature group comparison, this seems to point to the importance of absolute stature of the reference sample on the results. It is however possible that other factors do influence the results too. This will be discussed further in Chapter 7.

Not all fossils were equally significant for the tests. Especially the most human-sized one, OH 28 with 45.6 cm, was insensitive to most of the tests. The smallest one, KNM-ER 1463 with 31 cm, was the one with the most significant results. It is proposed that these patterns are seen because the results of the equations are clustered around the average of the modern human femur, and are therefore so close together that no differences can be discerned. KNM-ER 1463 however, is outside of the human femur range and is therefore placed at a position where the results for most modern humans will have to be extrapolated. This results in larger spread, as can clearly be seen in figure 5.13, allowing groups to be recognized as separated from each other.

The 110 equations present in the dataset forming the "wide selection" were selected into a "strict selection" for some of the tests. This was done to reduce the noise and double countings of the same reference sample. This would reduce the likelihood of false differences between groups to be recognized. Some tests were executed on both the strict and the wide selection. One of these tests had significant results for the wide selection and insignificant results for the strict selection. This was interpreted as a signal that selecting the dataset was useful, as no actual extra data were present within the wide selection it was likely that these significant results were derived from the double countings and hence falsely obtained higher intra-group coherence of some of the reference samples.

In the next chapter, three fossils will be estimated by the equations collected in the dataset. They will be compared with each other, and with an anatomically reconstructed stature where possible. This will hopefully give some assistence in understanding how to estimate *Homo erectus* stature.

Chapter 6

Hominin stature estimation

In this chapter, the collected equations will be applied to hominin fossils, in order to get a better grip on the differences of the results that they give. These test-cases will serve as illustrations to the discussion, that will be further be conducted in the next chapter.

Because the results of the previous chapter were somewhat ambiguous as to which factor of the reference sample was most influential on the result, an approach that is tried in this chapter is a matching on size between goal individual and reference sample and not on ethnic origin or environmental similarity.

How can we test the reliability and accuracy of estimation equations in general? Several criteria emerge from the literature. Performance of estimation equations in modern populations are tested by running the equations on individuals with known body size. The accuracy and bias of the method is calculated and expressed in mathematical constructs as the standard error (SE), standard error of the estimate (SEE). There are no extinct hominin individuals with actually known body size. There are individuals whose body size is likely to be more reliably estimated, because they are partially complete specimens.

Homo erectus partial skeletons that allow for a reliable anatomical stature reconstruction are unfortunately not known. KNM-ER 15000 is the one that comes closest, but his publishers still preferred to use regression to estimate his stature. *Australopithecus afarensis* AL 288-1, Lucy, is an example of such a fossil. This fossil is very instructive in the question of the estimation equation reliability. Geissmann (1986a) tested the results given by many estimation equations on the anatomically calculated stature (105 cm, taken as true stature). The mean of the results from 123 equations was 133 cm ± 8.2 cm, in a range of 105.9–145.6 cm. Only 7 of the equations came to results below 120 cm. The gross overestimation visible in these estimates, partly by equations used in this thesis as well, serves as a warning against putting faith in an average or clustering of answers. The parallel with the clustered results of this study (Chapter 5) is clear; the clustering of results from similar techniques does not increase the reliability of the results.

The equations used by Geissmann (1986a) were all based on simple linear regression, except for two major axis equations, and used modern human populations as reference samples. This similarity in approach is the reason for not accepting their agreement as reliable. When techniques that differ in approach agree on results, this is a reason for accepting their results as reliable. Different approaches use, for example, different input variables or different techniques. When both femur and tibia yield the same results, or when regression and anatomical reconstruction agree, this places confidence in both methods and in their consistent result. A method (set of equations based on the same population) that gives consistent results for several limb segments is likewise seen as reliable by Geissmann (1986a). If different limb segments show consistency in estimate, this is also an indication that the body proportions of the fossil individual were similar to the proportions of the reference sample because in those cases the elements all stand in the same proportion to stature as in the reference sample.

6.1 AL 288-1

Two fossils are available for anatomical stature reconstruction. The first, and to my knowledge uncontested, fossil is AL 288-1, Lucy, at 105 cm with a femur length of 28 cm (McHenry 1991a). It is not *Homo erectus*, not even *Homo*, but it is instructive to see which equations predict its stature correctly (see appendix). When statures between 100 and 110 cm are considered correct, 19 of 118 equations give correct results. These are:

Feldesman *et al.* (1990): Generic ratio, highest ratio;

Hens *et al.* (2000): Pygmy MA, Pygmy RMA, Pygmy ratio, Pygmy inverse;

Hens *et al.* (2000): Human classic, human ratio;

Jungers (1988): Bonobo OLS log;

Jungers (1988): Pygmy OLS raw, Pygmy MA raw, Pygmy RMA raw;

Olivier (1976): Pygmy correlation axis;

Ross and Manneschi (2011): Chilean males;

Feldesman and Lundy (1988): Indigenous South African male MA living height, female MA living height;

Feldesman and Lundy (1988): Indigenous South African female OSL skeletal height;

Sciulli and Giesen (1993): Indigenous American female skeletal height;

Dayal *et al.* (2008): Caucasian South-African females skeletal height.

From these 19 equations, 8 are based on Pygmies and three on other small-bodied Africans (these latter are counted amongst the "standard humans"). Furthermore, two out of three logged Pygmy equations and one additional non-logged Pygmy equation are between 110–112 cm. These are almost all Pygmy-based equations present in the dataset, although it must be remembered that all these equations are calculated on only two samples of Pygmies. Other equations generally give results which are too high. Remarkable is that the generic ratio gives a nearly spot-on result of 104.7 cm, while the highest ratio (African American-based) available from Feldesman *et al.* (1990) gives a result of 100.7 cm and the lowest (Eskimo-based) ratio comes to 110.02 cm. Therefore, the entire range of average population ratios given by Feldesman *et al.* (1990) gives a stature result for AL 288-1 which is within 6 cm of the anatomically reconstructed stature.

Three of these 19 "correct" equations are misleading. They calculate skeletal height instead of living stature; the results should still be adjusted with 10 cm according to Fully's method. The "standard human" equations which give correct results are the human OLS method of Hens *et al.* (2000) and the male Chilean equation by Ross and Manneschi (2011). Two more "standard human" equation give results of 111.3 cm (MA method by Hens *et al.* 2000) and 114.9 cm (Female OLS adjusted by Feldesman and Lundy 1988), while four hominoid-based, raw, equations are in the 95–99.99 or 110.01–115 cm range. Nonhuman hominoids especially have been expected to be of use in body size estimation of AL 288-1 (Aiello 1992a; Jungers 1988), for example because its femur length is close to the average femur length of bonobos and chimpanzees (29–30 cm, Hens *et al.* 2000).

In conclusion, results in the 95–115 cm range are given by equations based on hominoid, Pymgy and "standard human" based equations are not present amongst the results counted as "accurate", a situation similar to that described by Geissmann (1986a). In contrast, a relatively large amount of all Pygmy-based equations in the database yields results that are within 7 cm of the anatomically reconstructed stature. It is striking that almost all standard-human based techniques overestimate AL 288-1's stature, while almost all Pygmy-based techniques estimate its stature correctly. Apparently, smaller bodied modern humans are better models for small hominins, even if they are not exactly in each others size ranges. The good performance of the femur/stature ratio should also not go unnoticed. Not only did the generic ratio predict AL 288-1's stature spot-on, the most extreme ratio's present also performed on an acceptable level, both within 6 cm of 105 cm.

6.2 KNM-WT 15000

The second fossil allowing in some way for a comparison to its anatomically reconstructed stature is KNM-ER 15000 (femur 42.9 cm, tibia 37.5 cm; Ruff and Walker 1993). The original publication on its stature and mass came to a prediction of about 160 cm at death (Ruff and Walker 1993, 248), a result that was also achieved by Feldesman and Lundy (1988, 591). Ohman *et al.* (2002, 133) published a rather dissimilar stature-at-death of 141 or 147 cm, based on a modified Fully's method of anatomical stature reconstruction. The disagreement between these two groups of estimates lies in the assumption of similar allometry that is in the heart of both regression and ratio techniques; the assumption that *Homo erectus* body proportions were similar to those of modern humans. Ohman *et al.* (2002) however, reconstructed the length of the thoracolumbar column and came to the conclusion that it was small compared to modern humans with similar femur length and thus also small compared to KNM-ER 15000's own femur. Such a dissimilarity in proportions resulted in a shorter stature than when proportions would have been similar. This disagreement between published statures-at-death makes it difficult to compare the results at hand to "known stature". However, as Ohman *et al.* (2002) use a variant on the Fully technique, their results will be compared to this. This is not to say that these results are "true" stature.

The equations of the current dataset that give results between 140–150 cm for a femur length of 43.2 cm (maximum length in Ruff and Walker 1993, 247) are listed below.

Feldesman and Lundy (1988): Indigenous South-Africans female OLS, male OLS, female MA, male MA; all for skeletal height;

Sciulli and Giesen (1993): Indigenous Americans females and male; both for skeletal height;

Dayal *et al.* (2008): Caucasian South-Africans females and males; both for skeletal height.

All these equations calculate skeletal heigth instead of living stature. Their adjusted counterparts do not give results in the 140–150 cm range. All in all do none of the living stature equations give results in the range deemed correct following Ohman *et al.* (2002).

When results between 140–150 cm are seen as "correct" for KNM-WT 15000, no living stature estimation equations give a correct result. In complete contrast to AL 288-1, there are no Pygmy-based equations that give correct answers for KNM-WT 15000. They give higher results, in the region of 160 cm. This is also the case for the generic ratio. These results indicate that the proposed body proportions of KNM-WT 15000 by Ohman *et al.* (2002) are different from body proportions of regular-sized modern humans, small-sized modern humans and nonhuman apes.

The generic ratio equations predicts Lucy's stature spot-on, but reaches a stature of 160.3 cm for KNM-ER 15000, although (Feldesman 1992, 455) use a specific adolescent-based ratio to come to 157.4 cm. The strange result from this small comparison, and under the assumption that KNM-WT 15000 really was between 140–150 cm, is that AL 288-1 had a modernly proportioned femur to her total stature, while KNM-WT 15000 does not. Why is this the case? It is unexpected as AL 288-1 is much older than KNM-WT 15000, and if bipedal constraints forced this proportion, it would be expected to be present even more in KM-WT 15000 than in AL 288-1. Several explanations can be brought forward to explain this.

- The KNM-WT 15000 anatomical stature is less secure (true stature was actually larger than 140-150 cm).
- The individual was an exception (other hominins do have femur - stature proportions like modern humans), for example due to an abnormality, as proposed by Ohman *et al.* (2002).
- The individuals proportions are influenced by its peripuberal age, giving him temporarily deviating body proportions. The question is raised if juvenile individuals should be estimated by the use of adult-based estimation techniques (Porter 2002).
- Chance; e.g. body proportions in hominins varied more than expected, (bipedal) constraints are not very strong.

The comparability of body proportions is the key to correct body size estimation. When an individuals body proportions differ from those seen in modern humans, the number of equations giving correct answers decreases dramatically. When the first option is the case and actual stature was between 155–165 cm, then many more equations give correct results, extremely deviating results almost being limited to equations on nonhuman hominoids and log-transformed equations. The reference samples of which the femur average is close to the size of KNM-WT 15000 give results which are between 155 and 162 cm, these are the Trotter and Gleser (1952) Caucausian and African American female samples (160.8 and 158.3 cm respectively), the Genoves (1967) Mesoamerican male sample (161.5 cm), the Boldsen (1984) female Danish sample (160 cm), the Sciulli and Giesen (1993) female indigenous American sample (155.2 cm) and the Feldesman and Lundy (1988) small-bodied South-African female sample (157/ 157.6 cm).

New research that has come out (Haeusler *et al.* 2011) presented new bone (rib and vertebrae) material belonging to KNM-WT 15000, which is interpreted to contradict the assessment of abnormality of the individual (Haeusler *et al.* 2011, 580). This might influence the conclusions that can be drawn regarding body proportions, like Ohman *et al.* (2002) did, though Haeusler *et al.* (2011) do not enter this discussion.

6.3 D4167

6.3.1 Results of the dataset

This section will take the Dmanisi femur D4167 as a test-case to explore ways to come to a stature estimate for this fossil. D4167 (38.6 cm) is just inside the range of modern human femur lengths, as collected here. As example of the spread of femur length in a modern human population, the spread of the Dupertuis and Hadden (1951) female sample is between 38 and 49 cm for African Americans and between 34 and 48 cm for the Caucasian Americans. The average of the "standard human" groups, both female and male, however is for the largest part between 42 and 48 cm (see appendix). The D4167 value is close to the average of the Pygmy sample of Hens *et al.* (2000), which is 37.7 cm (average stature 140.5) and close to female average of the Mesoamerican sample of Genoves (1967), which is 39.6 cm (average stature 149.8 cm).

The estimates of the techniques employed here (the total or wide dataset) range from 82.9 to 220.9 cm, the mean estimate being 149.7 cm (SD 15.3, $N = 106$ formulae). The large SD would give that in the 68% confidence interval spread is between 134.4 and 165 cm, which is so large that it would hardly be informative biologically. The 95% confidence interval would have a spread of over 60 cm, giving quite irrelevant estimates. It must be stressed that these results given are calculated without their error ranges or confidence intervals, due to reasons of scope. The actual prediction range per estimate is thus likely to be much larger when a confidence interval is included.

Log-transformed and hominoid-based equations have in earlier tests (see Chapter 5) been shown to be prone to give extreme and outlier results, especially for femur lengths which are in the human range. Furthermore, the femur length of D4167 is not close to the average of the nonhuman hominoid species included here either. These averages are 34.6 cm (gorilla), 29.2 cm (chimpanzee) and 29.4 cm (bonobo), but as they are only used in combination or as exclusively bonobo sample the actual average of the reference samples will be close to 30 cm (see Hens *et al.* (2000); Jungers (1988)). When the log-transformed and hominoid-based equations are excluded to give $N = 82$, the estimate range reduces to 130.9 - 170.1 cm, with a mean of 150.1 cm ± 5.5 cm. The mean therefore is hardly affected, reflecting the large influence of the human-based equations in the first given average already, while the range and also the standard deviation are quite heavily reduced. The human-only equations thus give a 68% chance spread of 144.6–155.6 cm, a much more biologically relevant result. However, it is only valid when the hominin individual shared human body proportions.

As the Pygmy groups and the female Mesoamerican sample of Genoves (1967) have femur averages that are close to the value of D4167, the results of these groups are given an closer look. After all, these groups do not require extrapolation outside the range to calculate a

Table 6.1: Results for Pygmy groups and Mesoamerican females for D4167

Publication	**Reference sample**	**Result in cm**
Genoves (1967)	Mesoamerican females	149.7
Olivier (1976)	Pygmy correlation axis	144.09
Olivier (1976)	Pygmy regression axis	151.7
Jungers (1988)	Pygmy OLS	144.1
Jungers (1988)	Pygmy MA	144.7
Jungers (1988)	Pygmy RMA	144.4
Jungers (1988)	Pygmy OLS log	178.7
Jungers (1988)	Pygmy MA log	151.8
Jungers (1988)	Pygmy RMA log	151.2
Hens *et al.* (2000)	Pygmy inverse OLS	143.6
Hens *et al.* (2000)	Pygmy classic OLS	144.4
Hens *et al.* (2000)	Pygmy MA	144.4
Hens *et al.* (2000)	Pygmy RMA	144
Hens *et al.* (2000)	Pygmy ratio	144

result for this value. The Pygmy groups give the results presented in table 6.1.

The first thing to be noted is that all equations by Jungers (1988) and Hens *et al.* (2000) are based on the same sample, which is why only one of these equations is used for the statistical analyses in Chapter 5. The mathematical method again shows to be of little influence on the final result, at least in this case where extrapolation is not needed. The log-transformed equations do give deviant results from the raw data-based equations, one of which is especially unlikely (the OLS log result). If the results of the log-transformed equations are set aside, the picture that rises from these results is that of a stature between 143 and 152 cm, perhaps with a focus around 144 cm.

6.3.2 Stature estimates of the original publication

The original publication (Lordkipanidze *et al.* 2007) gave as estimates 134.3–155.7 cm (average stature without CI). These were based on the Trotter and Gleser (1952) Caucasian- and African-American males and females equations, and the Lundy and Feldesman (1987) African equations, see table 6.2.

Lordkipanidze *et al.* (2007) use several formulae by others to estimate the stature of the Dmanisi hominins. Those are using the juvenile humerus length (Ruff 2007), the adult humerus length, femur length, tibial length (Caucasian and African from Trotter and Gleser 1952; Lundy and Feldesman 1987) and the metatarsal I length (Byers *et al.* 1989). The

Table 6.2: Femur-based stature estimates for D4167 in Lordkipanidze *et al.* (2007)

Average in cm	CI in cm	publication	reference sample
154.5	145.3–163.7	TG 1952	Caucasian-American males
152.6	144.1–161.2	TG 1952	Caucasian-American females
155.7	144.4–166.9	TG 1952	African-American males
151.2	142.8–159.5	TG 1952	African-American females
138.5 (148.5)		LF 1987	African males
134.3 (144.3)		LF 1987	African females

spread of these six equations is large; from 134.3 to 166.9 cm, excluding the CI's of the equations from Lundy and Feldesman (1987) (they are not given). This spread is large in terms of biological relevance, see for example McHenry and Coffing (2000).

However, quoting (Ruff 2010, 169): "*Lordkipanidze 2007 did not correct their stature estimates derived from Lundy and Feldesman (1987) equations for soft tissue; this results in an addition of 10 cm to all estimates. Also, they used maximum femur length whereas Lundy and Feldesman (1987)'s equations are for physiological length, which appears from fig. 2 in* **?** *to be approximately 2 mm less than maximum length. Making these corrections results in a mean over all stature estimates in table S6 of 153 cm for the large Dmanisi individual.*" The corrections are the values between parentheses in table 6.2. This changes the picture to a more consistently stature estimate range. The spread without confidence intervals is similar to the 1 SD spread of the human-only results, and does also show similarities with the results of the like-sized groups (table 6.1).

Although the ratio method can be seen as a special, intercept-free, form of regression, it can be used here as an, alas only partially independent, different approach than the normal linear regression equations. The generic ratio gives 144.4 cm, on the lower side of the CI's of the Trotter and Gleser results, and very close to the adjusted Lundy and Feldesman results (144.3 and 148.5 cm). The lowest and highest ratio, which both gave very acceptable results for AL 288-1, gave results of 139.7 and 151.4 cm.

If the femur length of the reference samples of Trotter and Gleser (1958) are compared to the femur length of D4167, it can be seen that the difference is quite large. D4167 is 38.6 cm long while the Trotter and Gleser average of the right femur is 48.2 cm for African Americans and 47.1 cm for Caucasian Americans. As Trotter and Gleser are known to overestimate individuals smaller than their reference samples (as they themselves warn), stature for this individual is more likely to be accurately estimated with reference samples of similar femur length, which give a range of 143–152 cm. The generic ratio suggested that perhaps the lower side of this range would be expected for an individual exhibiting the general human femur to stature proportions.

6.4 Results from tibia-based equations

Three tibiae associated with femora are available for comparison of results, these are KNM-WT 15000, the Dmanisi large adult (D4167–D3901) and AL 288-1. The first two will be discussed here. Although the tibia-based equations should be evaluated just the way the femur-based equations have been evaluated in this thesis as they are subjected to the same problems as femur-based equations, they are used here as a cautious comparison with the femur-based results without such an evaluation.

KNM-WT 15000 has a tibia of 37.5 cm (Ruff and Walker 1993). A selection of formulae that use tibial length for stature estimation have been applied on this fossil. The results of these formulae are presented in table 6.3, again without confidence intervals. These results show a range from 156.7 to 178 cm, which is a relatively small spread compared to that of the femur-based results. Where the descriptive statistics of the human-only, no log, femur-based results give a $N = 82$ average of 161.2 cm, with a range between 139.8–175.2 cm and a standard deviation of 4.1 cm.

This smaller spread could very well be the result of the lower number of formulae used for the tibia ($N = 9$) in comparison with the number used for the femur ($N = 82$), but there also seems to be a different focal size for both bones. The tibia-based equations give taller statures than the femur-based equations. Remarkable especially is that the tallest result is given by a tibia-based equation, while the number of equations that is based on the tibia is so much smaller than the number that are based on the femur. Perhaps this is indicative of an elongated tibia to femur in comparison with the reference samples on which these equations were based.

The results for D3901 (30 cm) have a spread from 134.7 to 155.2 cm, a similar-sized absolute spread as KNM-WT 15000 though due to the size difference of 7.5 cm between the tibiae placed completely below the spread of KNM-WT 15000. Comparing these values to the femur-based results discussed above shows that the results of the Trotter and Gleser equations are very similar for both bones, between 151.1 and 155.7 cm combined. The other equations give results which are in this range too, except for an Nigerian-based equation (134.7 cm) and a Pygmy-based equation (144.3 cm). These results are in line with the femur-based results. As the Pygmy-based equation was preferred for the stature estimates due to the comparability of size, there is no reason to prefer another reference sample over the Pygmy sample for the tibia especially since the generic human ratio, the Pygmy tibia and the Pygmy femur equations all agree on a size around 144 cm.

Table 6.3: Tibia-based stature estimations KNM-WT 15000

Formula	**KNM-WT 15000**	**D3901**
TG 1958 African-American Male	167.5 cm	151.1 cm
TG 1958 Caucasian-American Male	178 cm	154.5 cm
Dupertuis1951 Male	171.4 cm	153.4 cm
Allbrook 1961 Kenyan Male	166.2	151.2 cm
Didia 2009 Nigerian Male	156.7	134.7 cm
Breitinger1937 German Male	170 cm	155.2 cm
Pearson1899 French Male	167.8	149.9 cm
Stevenson1929 Chinese Male	172.7 cm	150 cm
Olivier1976 Pygmy Male	158.2	144.3 cm

6.5 Conclusion

Although the total dataset has a large spread of results per fossil, this does not mean that this spread is random and equally divided over the whole range. A lot of the variation is on account of the hominoid-based and the log-transformed equations, because when these are excluded, the range of variation decreases. The remaining human-based equations give reasonably similar results, though Pygmy-based estimation method can stand apart, depending on the fossil that is estimated.

The anatomically constructed stature of AL 288-1 is treated as a given at 105 cm. Although close results are given by hominoid-, Pygmy- and standard human-based equations, most standard human equations overestimate the stature. The hominoid-based equations generally give quite good results, including some of the log-transformed, though some outliers are present. Good results from hominoid equations are not unexpected as their average femur lengths are close to the femur length of AL 288-1. It is however the Pygmy group that has consistently good results, although the many Pygmy equations are all based on only two reference samples. However, the good performance of other small-bodied samples is a support of the use of small-bodied samples for small-bodied goal individuals.

Another group that performed well is the ratio group. The generic human ratio gives a result which is correct within the centimetre, while the lowest and highest recorded human population ratio's give results which are only about 5 cm off. It seems that even hominins that are smaller than all living humans share the ratio of femur to stature, which might have something to do with locomotor efficiency. If even such a small individual, who has on other accounts such different body proportions (Aiello 1992a), shares this human feature, this might be present in other hominins too and might give practical credibility to the femur/stature ratio as stature estimator.

KNM-WT 15000's stature is less secure than that of AL 288-1. Although its publishers put it at around 160 cm at death (Ruff and Walker 1993, 248), an attempted anatomical reconstruction by Ohman *et al.* (2002, 133) came to 141 or 147 cm. None of the collected estimation equations confirmed this smaller stature. Instead, most agreed on a stature around 160 cm, including the generic ratio. New fossil material furthermore show that this individual might not have suffered from abnormalities as established by Ohman *et al.* (2002) which might result in a re-assessment of the body proportions and consequently re-assessment of the body stature estimates based on this.

The stature as estimated for D4167 by Lordkipanidze *et al.* (2007) is, adjusted by Ruff (2010), 144.3–155.7 cm (without confidence intervals). As the D4167 femur is very close in length to the average length of Pygmy femora, it is assumed that this reference sample is more suitable to the estimation of D4167 than the taller "standard human" reference samples as no extrapolation is necessary in the first case. The Pygmy results indicate a length of 143–152 cm, with perhaps a larger preference for the smaller part of that range. The generic ratio (not based on Pygmies), which correctly estimated AL 288-1's stature, resulted in 144.4 cm. All in all, it seems likely that the stature of the D4167 individual was around 145 cm, based on its femur length only.

However, several caveats must be remembered. First of all, the necessary application of confidence intervals will increase the range of values. Secondly, these results are only based on femur length, while more bones are available (Lordkipanidze *et al.* 2007). The femur is usually considered as the best predictor bone for stature in modern humans, but this does not mean that other sources of information for hominins should be ignored. The use of a Pygmy sample for a Dmanisi hominin further does not take into account any specialized adaptations both populations may have had. All of this is not taken into account here due to scope limitations.

The stature estimates from the tibiae of both KNM-WT 15000 and D4167–D3901 seem to agree with the stature estimates based on their femora. For D4167–D3901, equations based on the same reference sample (the Trotter and Gleser samples) give almost exactly the same results. This is also the case for the Pygmy sample, which was preferred because of similarities in femur size with the Dmanisi individual. The results from the tibia equations therefore give no reasons to change the estimated stature.

The tibia from KNM-WT 15000 gave broadly similar results to the femur. However, the spread of results was smaller, probably due to the much lower N, and the concentration of results seem to lie at a higher stature. If this pattern is confirmed by further research, this could support the notion that KNM-WT 15000 had body proportions different from those common in modern humans, perhaps indicating an adaptation. If distal elements are elongated in proportion to the proximal element, they would give higher stature estimates than the proximal elements.

This chapter in combination with the last two chapters represent three approaches to the estimation of body stature in hominins. Although these approaches and their results have been discussed in the chapters themselves, an attempt at an overview discussion will be done in the next chapter.

Chapter 7

Discussion

7.1 Doing body size estimation

The main question considered in this thesis is how reliable body size (stature) estimation methods are, especially for an accurate result for *Homo erectus* individuals, and which factors are most influential on the results they give. These questions are important because estimation methods are used to estimate the body size of hominin individuals, the results of which are subsequently used to understand the species' behaviour and biology better. Several approaches (Chapters 4, 5, 6) were taken to evaluate reliability from different angles. In this chapter, the results from these approaches will be discussed and placed in an overview framework. I will discuss the ways in which we can assess the reliability of estimation techniques when used to predict hominin body sizes. Though these questions give rise to a general theoretical discussion, the focus here is on the case of *Homo* (*erectus*) individuals.

Some of the questions posed in this thesis can be discussed with the results of the previous chapters at hand. As most methods of stature estimation are based on a reference sample, the quality of these samples were evaluated in Chapter 4. The study of 110 stature estimation equations in Chapter 5 helps to understand which factors influence the estimations in what way. As the predictor is kept constant at femur length, the two factors equations consist of are reference sample and mathematical technique. The analysis discussed the patterns that are visible in this data, and comes to the conclusion that of the tested factors, it is the reference sample that has the largest influence on the shape of the estimation regression equation. In Chapter 6 it was discussed in what ways three test-case fossils, AL 288-1, KNM-WT 15000 and Dmanisi 4167, could be given a stature estimate by comparing the results.

There are several methods to estimate body stature. The method that is most frequently used is the regression equation. Stature estimation equations are for the largest part simple linear formulae, which determine stature as predictor bone length * X and add an intercept to better fit the low and higher extremes of the samples. If the intercept is lost, the remaining formula is nothing more than the predictor bone length * the average ratio of bone length to stature of the sample, which is how the ratio formula works.

At the core of body size estimation regression equations lies the relationship of the predictor bone to the stature (Béguelin 2009, 2). There is variation in this ratio, also within a population (Trotter and Gleser 1951) and this is taken into account when the regression equation is build on the scatter of individual femur length and stature data. This is the reason that populations tend to have unique estimation equations. Variation in the predictor bone/stature ratios is an effect of differences in body proportions. Body proportions, and total stature, can be influenced by many factors. In Chapter 5, an attempt was made to determine for some of these factors how influential on the trajectories of the regression equations they are.

Factors that are influential in total body stature and possibly also in body proportions are genetic and environmental. It is known that stature is in part heritable. The heritability has been established around .8 (Hernández *et al.* 1998, 549), or for European men between .87 - .93, and for European women .68-.84 /.89-.93 (Silventoinen *et al.* 2003). It also has an environmental aspect, which can be ontogenetic in nature as a response to (lack of) nutrition and disease (Hernández *et al.* 1998, 549; Subramanian *et al.* 2010). However, stature and body proportions have also been thought to be influenced by environmental factors which work on a genetic level, as long-term adaptation to climate (Ruff 1991, 2002; Ruff and Walker 1993). Stature is therefore proposed to be determined by a complicated interplay between both genetic factors, which may be adapted to certain environmental factors, and ontogenetic circumstances that influence the individual in its possibilities of attaining the maximum genetically determined possible stature. Body proportions might be just so determined (Hauser *et al.* 2005).

As the body proportions of the reference sample determine the equation, the accuracy of a stature estimate is dependent on the similarity of the estimated individual and the reference sample. The method of the regression equation estimates the individual as if s/he was an average member of the reference sample (Hauser *et al.* 2005, 188). The matching of the goal individual to an appropriate reference sample is therefore paramount to an accurate result, but it also difficult. For the assessment of body proportions of *Homo erectus*, it would be necessary to have (semi)complete skeletons which are unfortunately quite scarce. Furthermore, it is not to be said that all *Homo erectus* individuals would have similar body proportions if such skeletons were available. To be able to fit a matching reference sample to any fossil hominin femur, it would be helpful to know what determines body proportions and stature exactly. If this would be known, and it could be assumed that *Homo erectus* body proportions and stature were determined likewise, it would be easier to argue which reference samples would be most appropriate. Therefore, an inquiry into some of the factors that might determine the outcome of estimation equations was executed in Chapter 5.

7.1.1 Comparing the three methods of stature estimations

The anatomical method

In the cases where it is possible to apply the anatomical method of Fully (1956); Fully and Pineau (1960), the results of this method are treated as more reliable and accurate than the results of other methods, e.g. by Feldesman and Lundy (1988); Formicola (1993). However, Fully's method is based on a reference sample and therefore should be evaluated for the quality of its reference sample just as the estimation equations are evaluated for this.

Fully's subjects were 102 concentration camp victims of WWII with mixed nationalities, this is the same sample that Olivier (1963); Olivier *et al.* (1978) used later. The usage of concentration camp subjects might be objected to, as they will have suffered deprivation perhaps already before entering the camp (upon which stature was measured) which could have effected stature (Porter 2002), and mass certainly. However, Olivier *et al.* (1978) shows that the measurements of the subjects fall between Rollet's data and that of Trotter and Gleser (1952), and so seem so be normal. These problems in addition to the uncertainty of the quality of the measuring procedures (footwear removal, inter-observer error control) (Porter 2002) would make this sample not to be preferred for yielding a stature estimation equation. Furthermore, at least Olivier *et al.* (1978) are silent on the ethical point of using data from Nazi victims from concentration camps.

These doubts are added to the point that a modern human reference sample can only give the situation for humans or more specifically that population, and will not give accurate results for extinct hominins if these hominins had a different distribution of soft tissue or another pattern of spinal curvature (Porter 2002). Even though Fully's method of stature estimation seems like the most accurate way stature can be estimated from bone material as it uses a larger part of the body, it is not definite that a result will be the spot-on correct answer.

With these problems regarding Fully's method in mind, what can be said about the estimations of the hominin skeleton, AL 288-1, best fitted for reconstruction with the anatomical method? The "anatomical method" used by Schmid (1983) in Geissmann (1986a) appears not to much to have been to use Fully's method, but rather to reconstruct stature by assembling the skeleton. Allowances for soft tissue have been made in this process, instead of using the allowances calculated by Fully on the reference sample mentioned after the addition of the bone heights.

The Fully techniques has been tested and revised by Raxter *et al.* (2006, 2007). By these authors it was found that the original technique gives results which are quite close to actual living stature and there is a high correlation coefficient, but it does underestimate the sample tested with an average of 2.4 cm. Adjustments made resulted in an anatomical method that was tested to give no directional bias and were accurate within 4.5 cm for 95% of the test population. When this method was applied, this was indicated with "adjusted Fully's method" in the appendix. This adjusted method seems to be preferable above the original method due to its explicit explanation of measurements, the mixed (African American and Caucasian American, and females and males) reference sample that was used and the accurate results it gives when tested to known-stature individuals. However, the results of the "old" Fully technique are not shown to be wholly inaccurate, and this makes it possible to use results from this technique with some faith still (Maijanen 2009; Raxter *et al.* 2006).

Regression and ratio

The difference between the "ratio" and "regression" equations evaluated in this thesis is the addition of an intercept in the regression equations. An intercept accommodates a better fit for the extremes of the sample, because it allows for a flatter slope. The ratio method is less sensitive and precise because of this. Its only instrument to fit an equation to a sample is the ratio itself while a regression equation is fitted by two instruments, the ratio and the intercept. As Pearson (1899) already explains, the ratio model can give very good results in the middle of a datacloud but usually fails at the edges. A solution proposed by Manouvrier (1892) was to separate the prediction equation in three: one for short people (left edge of datacloud), one for average people (middle of datacloud) and one for tall people (right edge of datacloud). A regression equations however can, because of the intercept, approach all data points in one equation. It is therefore simpler and more precise. Besides this, regression allows for the calculation of standard deviations and error estimates, while the simple application of a percentage does not.

The proponents of the ratio method in recent years have been Feldesman *et al.* (1989); Feldesman and Fountain (1996); Feldesman *et al.* (1990). Feldesman, Lundy and Kleckner (Feldesman *et al.* 1989) collected published data on a large sample (N=10,500) of modern humans from all over the world, computing the femur/stature ratio as 26.7% with a standard deviation of .55%. Lundy and Feldesman (1989), as reported in Feldesman *et al.* (1989), tested this ratio on repatriated Vietnam-era soldiers and came to more accurate height that the Trotter and Gleser (1952, 1958) regression equations, which were calculated on an American soldier sample. Feldesman *et al.* (1990) evaluate the ratio of the femur of the stature in 51 different populations of contemporary humans ($N = 12,149$). The average femur ratio of these samples was 26.74%, with significant racial differences but without significant sex differences. Africans are shown to be significantly different from the other ethnic groups (Caucasian, Asian), this result was repeated in Feldesman and Fountain (1996). It was also shown that using an estimate from another race was more inaccurate than using the general human ratio (Feldesman and Fountain 1996, 221). As extinct humans do not belong to any race, it was recommended to use the generic ratio for them. The spread of these 51 population average ratios is between 25.45% ("Eskimos") and 27.81% (US Black Military) and is normally distributed with "extremely low variability" (Feldesman *et al.* 1990).

Although this research into a shared human femur/stature ratio is stimulating, some comments can be made. The data used for calculation are population averages, thereby eliminating intra-population variation and reducing total variation. This affects the amount of error that can reasonably taken into account when using the femur/stature ratio. There are no standard deviations calculated with the calculation of a ratio, so no confidence intervals are known. An intuitive way of assessing probable spread of results would be to look at the smallest and largest ratio present, as has been done in Chapter 6. The averaging of populations reduces this spread of all ratios because it averages out the smallest and largest ratios present. Furthermore, because there is no indication of the total individual spread it is not known how large the ratio spread within a population is, e.g. if all individuals are very close to the population average or if the spread is as large as total human spread. This would be important information to assess probable error when using the femur/stature ratio, as a consistent large intra-population variation would indicate a need for a large confidence interval as the assumption that hominin intra-population variability was high is then better founded than the assumption that it was low.

The ratios presented by Feldesman *et al.* (1990) seem deceptively close to each other. After all, the smallest ratio is 25.45% while the largest is 27.81%, the average being 26.74%. However, they are tested to be significantly different (Ruff 1994). This can be illustrated by the use of figure 7.1: between the result of the highest and the average ratio lie all other equations, in this case the Caucasian and African American equations present in the dataset (NB. the smallest ratio is not indicated). However, even though the difference between the ratios is significant, the difference between the two results mentioned is only about 10 cm. The difference between the smallest and highest results, however, will consequently be higher. The results of these ratios at 46 cm femur length are 165, 172 and 180 cm (see appendix). This then comes down to the difference between actual stature and estimate one is willing to except. Taking into account current intraspecific variation in stature and the fact that the goal individual belongs to another species, I would be inclined to think that a 10 cm bandwidth would be acceptable as a smaller bandwidth would be unattainable (taking into account al the uncertainties of estimating a hominin). If a 15 cm bandwidth would be acceptable and still useful for further calculation on other aspects of *Homo erectus* biology is another matter, and is something that researchers doing these further calculations must address.

Regression is a better method that ratio mathematically (Hens *et al.* 2000; Konigsberg *et al.* 1998; Pearson 1899). The ratio method assumes isometry (the constancy of the fe-

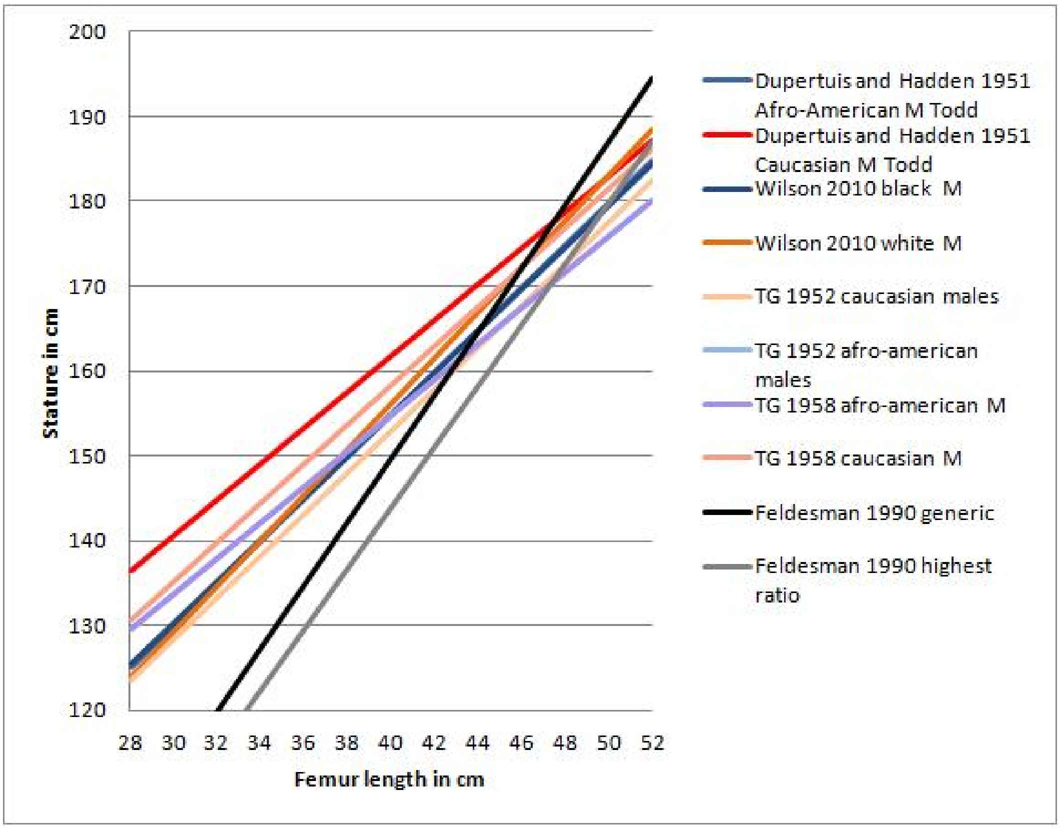


Figure 7.1: Several African and Caucasian American-based estimation equations

mur/stature ratio at all sizes of the femur and stature), while it has been shown that the femur/stature ratio is allometric (the ratio changes with absolute size) intrapopulationally in modern humans (Meadows-Jantz and Jantz 1999) (see below for a discussion on isometry). However, regressions are sample-specific and are widely documented to be unapplicable in other populations (Pearson 1899; Stevenson 1929) which causes great problems in hominin stature estimation as extinct hominins do not in principle belong to any modern human or nonhuman population.

The generic femur/stature ratio does not necessitates a choice between populations as it reconstructs a stature from a femur as if the individual was a “standard human”. This of course means that any individual which was strongly adapted away from isometry, either to climate, environment or some other factor, would be estimated with more error than when a regression estimation from a similarly adapted reference sample was used. However, the identification of the adaptation and the matching with an appropriate sample is a difficult thing. Regression equations have been shown to overestimate the stature of individuals with small femora, as was seen in the previous chapter with the estimation of AL 288-1 and in Geissmann (1986a). In the same comparison it was the femur/stature ratio which performed very well, with the lowest and highest ratio together creating a range of 11 cm around the “actual stature”. Therefore, the entire range of average population ratio’s given by Feldesman *et al.* (1990) gives a stature result for AL 288-1 which is within 6 cm of the anatomically reconstructed stature.

Even though regression is better founded in mathematics, the ratio method has been proven to be useful in all its simplicity. These methods can be used together, as a partially independent check.

7.1.2 Factors of influence in body stature estimation

The key to accurate body size estimation using regression is the similarity between the goal individual and the reference sample used. A regression equation is build on the data of femur length and stature. It is therefore the relationship of the length of the femur with the total stature that determines the fit of goal individual and reference sample (Béguelin 2009; Hauser *et al.* 2005).

However, this relationship is not isometric, not even within a population. This means that the proportion of femur to stature is in part also dependent on the absolute stature, as it changes with stature (Meadows-Jantz and Jantz 1999; Sylvester *et al.* 2008). The effect of this may be small within modern human populations (Sylvester *et al.* 2008) but its effect will likely be bigger when the range of stature is larger than that present within (standard) modern humans, i.e. in cases of extrapolation.

When a regression equation is built on a population, the allometry present in the dataset is incorporated in the formulae. When such a formula is used to predict a stature for a femur which size is outside of the range of the reference sample, the stature predicted is expected to follow the same femur/stature allometry. However, a short and preliminary look at the femur/stature ratio of Pygmies does not confirm to the allometric expectations as it is close to the average femur/stature ratio as confirmed by Feldesman *et al.* (1990). There is one Pygmy sample with data on the average stature and femur length present in the database, it is the sample that Hens *et al.* (2000) uses (see appendix). Its average stature is 140.5 cm, its average femur length 37.7 cm which gives a femur/stature ratio of 26.8%. The average human femur/stature ratio reported by Feldesman *et al.* (1990), who do not use a Pygmy sample to calculate this, is 26.7%. The lowest ratio comes from the "Eskimo" population, yielding a ratio of 25.45%, the highest from an African American military population of 27.8%. Pygmies thus display a very average femur/stature ratio and are therefore more "isometric" than "allometric" for femur length, and more so than expected based on the allometric scaling present in the "standard" American humans tested by Sylvester *et al.* (2008).

The effect of absolute stature on the results of estimation equations was tested by an ANOVA on different average size groups. This resulted in $p = .05$ significance for the smallest fossil and $p = .1$ significance for the largest fossil. The stature groups thus give significantly different results on the extremes of femur length, the smallest femur being out of the human range. Considering the isometry present between Pygmies and "standard humans", but also the fact that the population equations do take into account the allometry and other effects on the body proportions (e.g. climatic adaptation) present within the population (i.e. population-specific deviations from isometry) this would result in "standard human" population-specific equations giving results that increasingly deviate from isometry as the femur length used as input is increasingly away from the average femur length in the reference sample. The Pygmy-based equations would have a trajectory equally affected by within-population allometry and other factors, if present, but also have an isometric "grounding point" which is a much smaller femur length (average 37.7 cm instead of 45 cm) and therefore an estimation equation which is not close to the "standard human" equation, especially in the smaller femur range. As Pygmy average femur length is much closer to the smallest fossil femur in the test than the "standard human" femur length is, it is not surprising that the results of the Pygmy-based equations (so the results of the smallest stature group) are significantly different from the results of the other stature groups at this femur input point. This is illustrated in figure 7.2, which shows pygmy-based equations, modern American-based equations (both from Hens *et al.* (2000)) and the generic femur/stature ratio (Feldesman *et al.* 1990). It is further supported by the fact that it is the smallest size group that gives the most deviant values in comparison with the results of the other size groups, see table 5.11.

Average ratio of a population might be isometric (or the generic human ratio value), but it is only the average of the population while the estimation equation is built on the entire sample taken. The more scatter (the larger the spread of ratios in the sample), the more the estimation equation trajectory will deviate from that average (isometric) value, and the isometric trajectory belonging to it. This also explains the importance of the absolute stature of the samples. When the average femur/stature ratio of two samples are on the isometric value and their intra-population spread around this value forces the equation trajectory to a higher (or lower) but equal slope and intercept, the two equation lines will be similar in slope but parallel and with therefore always under- or over-predict and individual of the other population.

This procedure of matching size however does ignore the possibilities of influences other than allometry on the body proportions of both goal individual and reference sample. The "not-quite-isometric" relationship of femur and stature is probably not the only effect on the body proportions of an individual and thereby on the scatter around the average femur/stature value, and therefore not the only influence on the trajectory of the estimation equation. Its effect is described by Sylvester *et al.* (2008) as small and the insignificant differences for the more human-sized fossils on which the size groups are tested, supports this. Other factors of influence must be at work to explain the differences between like-sized reference samples.

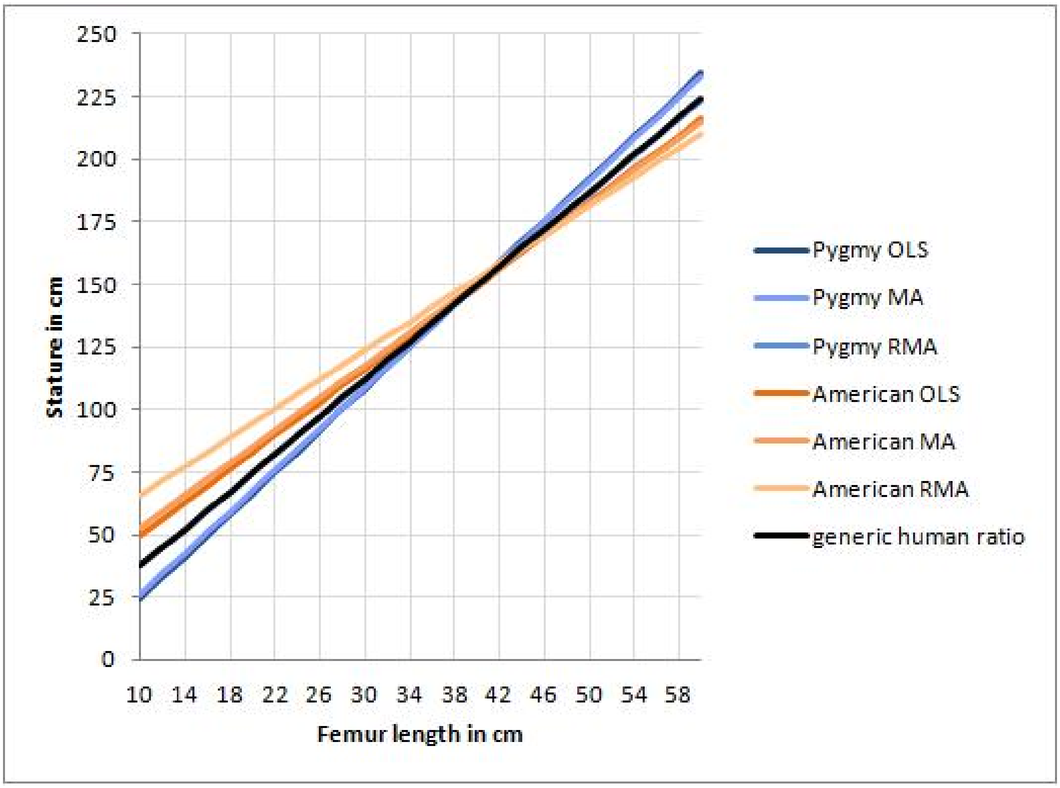


Figure 7.2: Several estimation equations based on Pygmies and Americans from Hens *et al.* (2000) and the generic femur/stature ratio from Feldesman *et al.* (1990)

Other hypotheses were suggested as to which factors would influence the results of stature estimation equations, one of them being the above mentioned total stature hypothesis. The hypotheses that were stated are the following.

- Adaptation to climate by changing of body proportions (Ruff 1994).
- Ethnic differences; different populations or "races" have different anatomical characteristics, for example due to genetic drift or founder effect.
- Differences in average stature between reference samples used; the influence of an allometric relationship between femur and stature.
- Differences in statistical technique used (ratio; OSL, RMA, MA; log transformations).
- Differences in results are due to random variation between human populations.

The differences in the statistical techniques used (OLS, MA, RMA) revealed to have no measurable statistical effect on the results of estimation equations that have been brought together in the dataset tested here. Equations that had been log-transformed did give other results than equations that were not log-transformed, but doubts remain about the exact influence and meaning as only very few equations were log-transformed. Hens *et al.* (2000) compare several estimator techniques and come to the conclusion that with univariate analyses there is no clear preference for a technique, as all five tested techniques seem to perform equally good (Hens *et al.* 2000, 776). The results presented here join that conclusion in so far as that these results show that none of the statistical techniques has such a large influence on the estimates that it is significantly different from the results of the other techniques. However, this concurrence might be due to the large part that Hens *et al.*'s equations are of the dataset on which the tests of this thesis were done. Unfortunately, only few other authors presented results of statistical techniques other than OLS so data was scarce.

The hypothesis of influence of ethnic background is grounded in the observance that regression equations of one population might perform badly on another population (Feldesman and Fountain 1996; Pearson 1899; Stevenson 1929; Trotter and Gleser 1952). This hypothesis was separated into two sub-hypotheses; the first was a classification based on the continent-of-origin of the reference sample, the second used a "specific geographic" origin of the reference sample. The last is a more refined classification, taking into account past migrations. To first explore a more basic division, the hypothesis first tested was that which stated that human and nonhuman-based equations would give different results. This indeed turned out to be the case, except for the smallest fossil. This deviating result was probably an effect of the large-scale extrapolation needed of the human-based equations in combination with no extrapolation at all needed of the nonhuman-based equations.

The classification on continent-of-origin is inspired by the many publications which give separate equations for "Black", "White", "Asian" groups, e.g. Dupertuis and Hadden (1951); Feldesman and Fountain (1996); Trotter and Gleser (1952, 1958). These classifications on the commonly recognised and long-held ideas of the "races" are essentially continent-of-origin based. As the authors (e.g. Feldesman and Fountain (1996); Trotter and Gleser (1952)) which have made such distinctions on ethnic background do not control for other factors, such as socio-economic background, environment or climate, it is assumed by me that it is genetic background (e.g. via founder effect/genetic drift) which is supposed to control the different body proportions of these populations in this hypothesis. For if climatic adaptation would be seen as the influential factor, the grouping factor would not so much be racial affiliation but e.g. average temperature of area of provenance. The racial classification that is taken is so broad environmentally that it incorporates several of the large climatic zones within every one of these races.

This view is probably a result of history and of American research focus. There is a fundamental difference visible in North America as a continent and the Old World. All populations except for the group of indigenous populations are in North America because of large-scale migration in the recent past. They were all displaced from their original place of origin, thereby removing any direct link with their supposed environment-of-adaptation and perhaps immediately starting adaptation to the new environment. Next to that, the migrated populations mixed with other populations, which made it impossible to separate populations on any level more precise than continent-of-origin. The recognition of this last level being facilitated by racial segregation until just a few decades ago in the USA. This racial level is therefore the only way of devising population-specific stature estimation equations for most North-Americans, as recommended following research (Pearson 1899; Stevenson 1929).

This is in contrast to the situation in the Old World (and some indigenous Americans), where the populations used are identified at the level of nationality (e.g. French (Rollet 1888), Polish (Hauser *et al.* 2005)). These European country-level populations usually all inhabit the same climate zone and it is a reasonable assumption that the population has lived there at least for some centuries and is reasonably homogeneous, albeit obviously being influenced by migration streams from other countries, and also from other continents in the recent years. The same could be the case for African and Asian populations, but no information on this is present in the current dataset.

Even if it is the only level of origin American populations can be distinguished on, the question remains how valid the distinction between the three races is. The existence of a difference between the equations based on African and Caucasian Americans in the samples of Trotter and Gleser (1952, 1958) would indicate that these (sub)populations at least harbour some differences, which would probably be due to genetic differences as all these individuals grew up in the USA. Socio-economic differences could also have an effect, but information on the individuals' background is unavailable so cannot be assessed.

The statistical tests executed in this thesis on the continent-of-origin classification indicated that this grouping was not completely irrelevant to the results. However, a direct comparison of all four groups resulted in no significance for the $p = .05$ level. However, there are significant differences reported between the results of the European and African origin group for the two smaller fossils.

Research by Feldesman and Fountain (1996) on the similarity of the femur/stature ratios of populations grouped into three large ethnic groups came to the conclusion that there was a significant difference between their Black and White groups, and between their Black and Asian groups but not between their White and Asian (including Indigenous Americans) groups. A large within-group spread was present here. These results are not contrary to the situation encountered in this thesis, although the tests are not completely comparable as some other tests were executed. The ANOVA on all four groups in the test in this thesis was insignificant, which corresponds to the lacking difference between some of the groups of Feldesman and Fountain (1996), and the European and African origin group compared did result in differences for some fossils, corresponding to the difference found between the White and Black group of Feldesman and Fountain (1996). However, a similar result should not be unexpected, as both datasets show much overlap in the reference samples used.

Feldesman and Fountain's conclusion from statistical research was that while the correct racial ratio or equation gives more accurate results for a person from that ethnic group than when the generic femur/stature ratio is used, it gives less accurate results for a person from another ethnic group. As extinct hominins by definition do not belong to a modern human ethnic group, they recommend the usage of the generic ratio on these hominins.

However, Holliday and Ruff (1997) explained the wide intragroup spread of Feldesman and Fountain's data by its large geographical and climatic spread (the climatic hypothesis). While the research of Feldesman and Fountain (1996) focussed on ethnic origin, it ignored the, in the eyes of Holliday and Ruff (1997), important factor which is body proportions. These body proportions are proposed to be somewhat related to ethnic origin due to shared ecogeographical conditions to which body proportions are adapted (Holliday and Ruff 1997). Because although all of the three races encompass several climatic zones, these are not the same climatic zones. According to them, this ecogeography is the condition that should be matched between the reference sample and goal individual. Testing this ecogeographic hypothesis was however impossible in the scope of this thesis, as a large part of the current dataset consists of American samples with the problems of relating them to a climatic zone as described above.

A more subtle geographic grouping was devised to take into account information on migration and true ethnic affiliation. This classification based on a more specific geographic origin of the reference samples did give quite significant results for some of the tested fossils. It is therefore concluded that the specific geographic origin does have influence on the results. This conclusion however, is only valid when the Pygmy group is included in the analysis. The reason for the Pygmy group to be so different is unknown, but it might be related to the issue of allometry discussed above. The Pygmy group would then be different because they are so much smaller in absolute stature than the other human groups. However, It could also be due to bad grouping criteria or other data issues.

As stated, the climate adaptation hypothesis could not be tested with the data at hand. If the ethnic origin hypothesis give significant results in this research, is this evidence for the rejection of the climate hypothesis? It is not, as there can exist overlap in the groups formed on "ethnic" criteria and groups formed on climatic criteria. An example of this is the European group. All reference samples from Europe group together as "race", but they also have a restricted geographical provenance of which the climate is for a large part labelled as "temperate". And because West-European samples make up a large part of the entire Caucasian group because a lot of research as been done in this part of Europe (e.g., only one Slavic sample is included). Relatively many equations exist for nationalities of the relatively small countries of Europe.

Furthermore, the hypothesis tested with the specific geographic classification is not strictly the ethnic hypothesis as it is stated above. The groups tested by Feldesman and Fountain (1996) were three "races" '('Black", "White", "Asian"), more similar to the continent-of-origin classification here. In an attempt to make this continent-of-origin classification more subtle by incorporating differences on ethnic affiliation (e.g. excluding Egyptians from the "Africa" group) and migration history (e.g. separate immigrated Americans from their groups of origin), the possibilities of environmental influence are larger. The specific geographic groups can therefore been seen as grouped sensitive to both ethnic as geographical characteristics.

To purely test the climatic hypothesis, it would be necessary to classify each reference sample to a climatic zone and then compare these climatic groups with each other. This still requires that population history is known, for a population should have had the time to adapt to the climatic circumstances. The climatic influence on stature is by Ruff (2002) clearly seen as a true genetic adaptation, and not as an ontogenetic development. Ruff (1991, 1994a, 2002) presents several approaches to the classification of climatic zones in order to be able to test the climatic hypothesis, which come down to classification on latitude for example by comparing lower-latitude groups with high-latitude groups (Ruff 1994). These comparisons do seem to show different patterns, but for example the correlation between stature and latitude is only .23 (Ruff 1994, 76). Body proportions and body mass however, might be just as well be influenced by climatic adaptation without a large difference in stature (Ruff 1994; Ruff and Walker 1993). This would influence stature estimation through the affected body proportions. A test that points to a small influence of climatic adaptation is that of Kurki *et al.* (2008), which indicated that in regions that do not experience climatic extremes other factors than climatic adaptation (e.g. energetic efficiency) have influence on body size and proportions too. This might be something to take into account, though it remains to be seen how this can be effectuated.

So what can be said about the influential factors on body proportions and stature? The most influential of the factors tested seems to be the reference sample. Which aspect of the reference sample it is that is most influential is however difficult to determine. It is difficult to separate the environmental adaptation and the non-climatically influenced genetic influence, as climatic adaptation is proposed to work via genetics. The climatic zones are represented in this dataset mostly by populations which have a similar ethnic background, as most groups are either originally from Europe or from Africa. A comparison between continent-of-origin groups "Europeans" and "Africans" might therefore be very similar to the comparison between "temperate" and "hot". This is due to the limitations of the dataset.

Another aspect of the reference sample that seems to be important is that of the absolute stature. It has been shown that extrapolating outside the size range of the sample indeed does often overestimate (see Chapter 6), and in Chapter 5 significant differences between the results of stature-based groups were present in the most extremely sized fossils (the smallest femur especially). Furthermore, the in- or exclusion of the Pygmy sample (a geographic group defined on stature) seemed to influence the "specific geographic" results the most of all groups included. It therefore seems wise to choose a reference sample which is close in size to goal individual, for example by comparing femur lengths.

7.1.3 Research problems

The research executed here is limited in scope and data. The total N, especially when divided over groups, is low. The total dataset of collected equations is larger than the set that can actually be used for most of the tests. For example, most of the equations come in pairs, as a female and male equation for the same population. Several of the reference samples have been used for more than one set of equations. Including all of these equations would inflate the importance that that specific reference sample has in the set. In effect, such samples would be counted several times. This will result in inflated levels of intra-group cohesion. As this is used to assess the goodness-of-fit of the grouping variable used, these double countings were avoided.

Some geographic origin areas are under-represented, only a few groups have a decent N. This is not only due to the selection of equations that is presented here, as effort has been made to include a range as wide as possible. More research has been done and more regression equations have been made on groups from certain backgrounds, especially Europeans and non-indigenous Americans. Information on reference samples from Africa, Americas, Asia and Australia is scarce; only a few samples represent complete and in environmental aspects diverse continents. This prohibits effective research into the exact influence of environmental circumstances on stature, and might have consequences for the availability of stature estimation methods which would be theoretically best.

A very large problem is the under-representation of Africans and Asians. Of the African group, only small-bodied reference samples have been included, and they have been divided into two groups: Pygmies and other Africans. The Pygmies are treated as a separate sample because they are defined on their stature as being different from all other human groups. The other Africans are in this dataset only represented by the small-bodied samples of Feldesman and Lundy (1988). This leaves the largest part of the continent and of the tropical body proportions (Ruff 1994) unrepresented. This omission in the dataset results in the "African origin" (continent) group for a large part being formed by Americans from African descent, and the African group (except Pygmies) being completely unrepresented in the tests on the specific geography groups. This same problem applies to the Asian group, they are missing from every analysis as there are hardly any data available at all.

7.2 *Homo erectus*

Palaeoantropologists try to understand the patterns in hominin evolution and the biology of the species in this lineage. It is thought that knowledge on the body size of the species will help elucidate other aspects of the biology and evolution. Body size, both stature and mass, are thought to be a central aspect of a species' biology due to the correlation with so many other variables, be it anatomical, physiological, life history, or behavioural (Damuth and MacFadden 1990; Foley 1987). This makes body size an often studied characteristic and starting point for palaeontologists to infer the aspects of a species which cannot be observed in the fossil record.

Body size has been used in hominin research to make inferences about the environment and life style of species and signal evolutionary changes in this. Much emphasis has been placed on the difference in body size between the australopithecines and *Homo erectus* or *Homo ergaster*, which was by McHenry and Coffing (2000) presented as the change between the australopithecines including *Homo/Australopithecus habilis* and *Homo rudolfensis*. This change was quantified as a difference of 23 kg and 29 cm for males and 19 kg and 50 cm for females (McHenry and Coffing 2000, 127).

The general benefits associated with a larger body size are of competitive origin: increased predatory efficiency and avoidance, increased thermoregulatory and metabolic efficiency, increased intelligence and longevity (Foley 1987, 120). However, a large body also requires more absolute energy for maintenance and growth, and can only be sustained when sufficient food is dependably available.

The larger body size of the species of *Homo* other than *Homo habilis* have been seen as a part of a changing pattern of ecology, diet, home range size (Antón *et al.* 2002), locomotor efficiency (Steudel-Numbers 2006), thermoregulation (Ruff 1991) and energy expenditure (Aiello and Key 2002; Aiello and Wells 2002). It has been used to explain the rate of dispersal of *Homo erectus* to other continents (Antón *et al.* 2002), but also to suggest changes in reproductive strategy and cooperation (Aiello and Key 2002).

The image of *Homo erectus s.l.* that has been created by these publications is one of increasingly more human hominins, living a life that is more like human hunter-gatherers than like chimpanzees or bonobos, or the australopithecines (Wood and Collard 1999). This is not only founded in a perceived large body size, but also in anatomical characteristics interpreted as evidence for obligatory terrestrial bipedalism, a diet which required similar mechanical properties as that of *Homo sapiens* and a more human-like developmental pattern (Wood and Collard 1999, 202–203). Furthermore, *Homo erectus* has been found outside of Africa, unlike earlier hominins. It seems to have dispersed widely in a relatively short time, as it has been found in Africa, the Caucasus and the far East in a geologically similar time frame Antón *et al.* (2002). The species is also associated with technological innovations (Acheulean technology), but certainly not at all sites and all times (de Lumley *et al.* 2005).

However, for body size to play a role in these kinds of complicated evolutionary analysis, it must be reliably reconstructed. I hope to have shown in this thesis that the accurate estimation of hominin body size is dependent on the fit of reference sample with the goal individual, not only matching in body proportions but also in general size. AL 288-1 matched in femur/stature ratio with modern humans, but most of the human-based estimation equations still consistently overestimate her stature. Samples which are closer to her in femur size (and stature) give estimates which are closer to the anatomically reconstructed stature.

The larger postcranial bones that appear in the fossil record from 1.9 mya onwards Antón (2003), will have belonged to taller individuals than the smaller bones. These larger femora have sizes that are close to or in the human femur size range. This in combination with the close phylogentic connection with modern humans would seem to make a human reference sample less inappropriate for these bones than for for bones outside of the modern human femur range. The modern human femur range lies, from the information gathered here (see appendix), between 34 and 54 cm, although the entire range is not known for many of the reference samples. Many *Homo* fossils are therefore in or just below the human femur size range (see appendix), which would enable the matching of fossil femur and reference sample according to femur size.

Some of the groups of equations tested for differences in Chapter 5 did show significant differences, which was interpreted as a signal that the classification was biologically valid. However, most of these significant differences were only attained in the fossils outside or just at the human femur size range. The fossil of average human size, OH 28 in this case, was a lot less sensitive to grouping differences. This was explained by the concentration of results of all modern human groups for the size range of their reference sample, at around 45 cm femur length. If this interpretation is correct and the differences between the groups are not significant in the human size range, this would mean that the differences between these groups are also not significant for fossils in the modern human size range. The choice of reference sample then would be of less importance for these fossils than for fossils which are out of the human range. However, questions about the appropriateness of the dataset collected and used in this thesis (see above) still apply, and the conclusions that can be drawn from them are by no means certainties.

Size considerations have not been the only criteria on which fossils and reference samples have been matched. The climatic zone and environment to which the fossil was thought to be adapted have been used as arguments to use modern human reference samples from that climatic zone or environment for that fossil (Ruff and Walker 1993). For East African *Homo erectus* it has been implied that its adaptations must have been for a hot, dry and open environment, which would have advocated a tall stature but especially small body breadth (bi-iliac stature) and disproportionally large limbs-to-stature thereby increasing surface-to-mass proportion facilitating the cooling of the body in such an environment Ruff and Walker (1993).

It has been attempted in this thesis to evaluate the body statures of two fossils which have, amongst other species, been assigned to *Homo erectus*: KNM-WT 15000 and D4167. Both femora have sizes which are in the human range, 43.2 cm (KNM-WT 15000) and 38 cm (D4167). The first is properly human-sized and close to the femur average of several populations, most of them the female average. For example, the Trotter and Gleser (1952) Caucausian and African American female samples (resp. 43 and 43.7 cm), the Genoves (1967) Mesoamerican male sample (43.2 cm), the Boldsen (1984) female Danish sampe (42.5 cm) but also close to the African samples of Feldesman and Lundy (1988) at 42.3 cm (females). The second is around the average of the Pygmy groups present in the dataset.

The stature of KNM-WT 15000 has been anatomically reconstructed by Ohman *et al.* (2002) but the results of this reconstruction were not in agreement with any of the results of the estimation equations or the femur/stature ratio. This at least raises some questions regarding the accuracy of the anatomical stature reconstruction, as it would be expected that if AL 288-1 can be estimated more or less accurately with some of the equations, this individual with a more human-sized femur would be estimated at least equally good.

KNM-WT 15000 and D4167 have been assigned to the same species and are only .3 my apart from each other in time (Lordkipanidze *et al.* 2007; Ruff and Walker 1993). They show a difference of 15 cm between the results that seem most realistic. Such a difference is not unseen variation within a current populations, let alone the entire human species, but it is too large a difference to see these two individuals as “similarly sized”. Furthermore, KNM-WT 15000 was not even fully-grown yet, although this remaining growth might not have been that large (Graves *et al.* 2010). On the other hand does D4167 represent the “large adult” of the Dmanisi palaeodeme (Lordkipanidze *et al.* 2007), even though the “adolescent” individual is already predicted to be larger than this adult individual.

It might very well be that within *Homo erectus* stature is variable and flexible, like it is in modern populations. Though stature has a very strong genetic component (Silventoinen *et al.* 2003), it is also responsive to environmental factors, which determine if the genetically determined maximum stature is achieved during growth. These very personal influences on stature are very likely to have had their effect on *Homo erectus* individuals as well. It is also possible that populations of *Homo erectus* have adapted to local circumstances, for example the climate and humidity. The scale of time and therefore possibilities for adaptation is of course much larger in *Homo erectus* than in the modern human populations. It is therefore no surprise that within sites with contemporaneous *Homo erectus* the crania tend to cluster on morphological characters with each other rather than with crania from other, less contemporaneous, sites (Gilbert 2008, 356). When the crania from morphologically distinctive groups of *Homo erectus* than a quite flexible trait like stature might be expected to show variation between populations of *Homo erectus*.

Finds from the last years indicate that this variation might have indeed been quite large. The difference between KNM-WT 15000 and D4167 might be just a little more than between the averages of the sexes in modern populations (the difference between average stature of males and females in the dataset is between 10 and 15 cm), the range of hominin stature in the Pleistocene is larger than what these fossils represent. If femur length is taken as a proxy for stature, it is clear than femora sized like average (and large) modern human femora are present (OH 28, KNM-ER 1808, Trinil I-III) but also femora which are below the human size range. Exemplary for a small-statured *Homo erectus* is the Gona pelvis dated between 1.4–0.9 mya, with an estimated femur length of 32.1–37.2 cm (Simpson *et al.* 2008). The average of this length, 34.6 cm, is almost out of the modern human range as represented by the dataset in the appendix. Only one population has a female femur range that encompasses this value. The accompanying stature estimate was 123–146 cm (Simpson *et al.* 2008, 24S). Such postcranials from small-bodied individuals are reflected in the finds of small *Homo erectus* crania, like KNM-OL 45500 (Potts *et al.* 2004) and KNM-ER 42700 (Spoor *et al.* 2007). These crania show that the range in variation in cranial size and morphology in *Homo erectus* is also larger than previously thought.

Although not everybody agrees (Ruff 2010), these finds seem to indicate the the variation in body size and body shape in *Homo erectus* were larger than hitherto thought. A very important signal that especially the Gona pelvis seems to give is that the climatic adaptations seen in modern humans do not apply to *Homo erectus*, as this is an individual from a low latitude and supposedly warm environment with a very broad, instead of narrow, pelvis (Simpson *et al.* 2008, 1091). If indeed such adaptations are not present in *Homo erectus*, this will have consequences for its stature estimation. This would mean that the choice of a population on the grounds of shared climatic adaptations is no longer a consideration, and other criteria need to be used to match a reference sample to the goal individual. The implications of this variability in body size on the aforementioned biological and behavioural inferences made on a larger body size of *Homo erectus* are as of yet unclear and this perhaps promises the possibility of new and exiting developments.

7.3 Possibilities for improvement

Are there ways to improve the estimation methods that have been evaluated in this thesis? The regression equations for this research have been limited to equations that only use femur length as a predictor. However, multiple regression on several variables is seen as a more accurate method than a when a single predictor variable is used (Damuth and MacFadden 1990). The addition of tibia length or the width of the femur are seen as variables that add accuracy and reliability (Porter 1999, 110).

The material used for taxonomic classifications for hominins is mostly cranial. Postcranial material that is unassociated with cranial material is hard to assign to a species. When postcranial material is used for stature estimation the resulting estimates are difficult to associate with a species too. It would be helpful if cranial material could be used for body size estimation as well. This has been done for body mass estimation (Aiello and Wood 1994; Kappelman 1996; Spocter and Manger 2007), and for stature on modern humans (Pelin *et al.* 2010; Ryan and Bidmos 2007). Although this last article did not show very optimistic results for the possibilities of calculating stature from cranial traits, it would be interesting to assess the possibilities for body size estimation on cranial material further. However, it is likely to be quite difficult due to the dependency of many cranial features on total cranial size (Antón *et al.* 2007).

The knowledge on hominin body proportions and body size can only benefit by the addition of more sources of information on these characteristics. Next to cranial material there are other sources of information, like footsteps Bennett *et al.* (2009) and of course other bones like the tibia and vertebrae for which modern human formulae are available (e.g. Didia *et al.* (2009); Duyar and Penil (2003)). Such equations can also help evaluating the different body proportions.

More reference samples from specific environments (Kurki *et al.* 2008, 2010), preferably homogenous and with a hunter-gatherer mode of subsistance are needed. Homogenous populations specific to an environment are nowadays perhaps difficult to find. A solution that has been used by e.g. Maijanen and Niskanen (2010); Sciulli and Giesen (1993) is to use archaeological samples, under the assumption that these samples have been less influenced by large-scale admixture. For similar analyses as in this thesis it would be a valuable addition if more populations from Africa (tropical and non-tropical) and Asia could be used, instead of populations that have an origin in Africa several generations ago. Such data collection might be aided by new techniques, like radiographic or MRI-scanning (Kieffer 2010), which can overcome the difficulty of obtaining stature measurements during life and bone measurements after death.

In this thesis, the "body size" that was focussed on was stature. Stature has its benefits over body mass, for example its relative stability in a person once adulthood has been achieved, while body mass can be quite variable. However, calculations for e.g. energetics are made using estimates for body mass, so an evaluation of the accuracy of body mass estimates is not useless. Unlike stature estimation equations, body mass estimation has not seen much interest from forensic scientists and therefore the number of equations available and the number of reference samples used is a lot lower. The extension of research and data on body mass would be very useful.

Chapter 8

Conclusion

8.1 Results from the research

Average body size in the form of stature and mass is a very useful characteristic of a species to know when one is interested in this species' biology and behaviour. The body size of a species is correlated to many of the biological characteristics which are unknowable for extinct species, such as diet, except for approximation through other, known, characteristics. Body size is one of the known characteristics which are used for such approximation techniques. In *Homo erectus* research, for example, this can be seen in Aiello and Key (2002); Antón *et al.* (2002); McHenry (1994); McHenry and Coffing (2000).

When the predicted body size of *Homo erectus* is used for such further research, it is essential for the predictions to be correct. This thesis has been discussing the techniques that are used in body stature estimation in regard to their usefulness, reliability and suitability for *Homo erectus*. For the estimation of stature in modern humans, many techniques are available. It has been recommended to use the technique that is most suitable for the individual at hand, which in humans translates to using a estimation equation that is specific to the population and to the sex to which the individual belongs (Hauser *et al.* 2005; Pearson 1899; Stevenson 1929; Trotter and Gleser 1952, 1958). For extinct hominins, the choice is not so simple as they do not belong to any of the current populations. So how to know which of the estimation techniques is most suitable?

Three approaches to evaluate the reliability and suitability of stature estimation techniques for extinct hominin have been used in this thesis. For this purpose, 110 stature estimation equations were collected. Most of these have been formed on modern human reference samples, but also nonhuman hominoid-based equations were included. They all used femur length as the predictor variable. These equations were used for the calculation of stature estimates for 25 hominin fossils with a range of femur lengths. These estimates are presented in the appendix. Due to the scope of the research, it was not possible to calculate the confidence intervals alongside these estimates.

The first procedure (Chapter 4) for evaluation of the equations looked at the quality of the estimation techniques themselves. This would also identify techniques which would be less reliable to use on modern populations, as it looked at quality characteristics as the correlation coefficient and the number of individuals in the sample. There are flaws to find in any reference sample. The samples that are most optimized on the criteria stated, are the male military Caucasian and African American samples of Trotter and Gleser (1952, 1958), although some doubts still remain (see also Porter 2002). Other samples have problems due to low N or the usage of limb measurements instead of femur bone measurements. However, some of such samples have other characteristics that can be valuable in hominin size estimation. For example, Pygmy samples extend the range of human stature and femur length, which might be beneficial when hominins smaller than the average human need to be estimated.

The second procedure (Chapter 5) used statistical tests on the differences between groups of estimation equations to try to identify the patterns that cause the differences between the estimation techniques' results. After all, there must be a reason why all these techniques yield different answers. If these influences are known, the most suitable sample for *Homo erectus* will be easier to identify. There are several factors that are of possible influence on this, like the predictor used (in this thesis only femur length is taken into account), the mathematical technique used, and the reference sample used. The factor "reference sample" is the most likely to be of large influence as can be seen from the warnings several authors have given regarding exclusively using an equation on the population on which is was made (Hauser *et al.* 2005; Pearson 1899; Stevenson *et al.* 2009). Several hypotheses are given that could explain why and how differences between reference samples could have arisen. The first explanation is the possible influence from the average stature of the reference sample. The second hypothesis is the ethnic origin of the samples, in which samples that share ancestry give more similar estimates probably due to shared genetics (based on Feldesman and Fountain 1996). The third hypothesis posits that populations give similar results because they are adapted to the same climate (Ruff 2002, 1994). The latter could not be tested here due to unavailability of the necessary data.

The second procedure divides the equations into groups according to the classification belonging to the hypothesis to be tested. In the case of mathematical technique, three groups are formed. One group consists of equations that are calculated using the OLS method, one group of those calculated using the MA method, and one group of those calculated using the RMA method. When these groups are compared using ANOVA in IBM SPSS 19 and a significant difference is present, i.e. that these groups are so different that they could not have come from the same group, it can be concluded that indeed mathematical method is a factor of influence on the final estimation result.

From these analyses, a mixed picture arises. The mathematical method does not seem to have influence on the estimation results, as no significant results are attained. The reference sample classifications do give significant results for several tests, but also some insignificant results. The significance seems to be related to the size of the fossil that is tested: OH 28, of average human size, seems to be more insensitive to the different classifications than, especially, the smallest fossil tested (KNM-ER 1463). This is the case when continent-of-origin ("racial") ethnic groups are tested and also when a more specific geographic classification is used. The latter classification, a more refined version of the first, does give more significant results than the first. The group with the largest influence in these classifications seems to be the Pygmy group. This seems to be in line with the significant result attained in the smallest fossil when four stature groups are compared, as the Pygmy group also forms the smallest stature group. It argues for the usage of a similarly-sized reference sample. Otherwise, clearly defined groups of populations (African and Caucasian Americans) did yield significant differences too. Absolute stature therefore is not the only influence on the results. It however was not possible what exactly caused the other differences, be it general relatednes, climatic thermoregulatory adaptation or some other adaptation (Kurki *et al.* 2008; Ruff 2002, 1994).

The third procedure uses the information gathered in the other two approaches to look at specific fossils. Especially fossils which have an independent stature estimate can act as a control to the estimates of the regression techniques. An independent stature estimate must rely on other data and/or techniques. There is only one method that is separated from the above mentioned estimation techniques, this is the anatomical stature reconstruction method. This method, known as Fully's method (Fully 1956; Fully and Pineau 1960; Raxter *et al.* 2006) calculates the stature by adding up the lengths of all the bones in the body that attribute to stature. It will be clear why this method can unfortunately be used only rarely in palaeoanthropology, although the results are highly accurate (Maijanen 2009; Raxter *et al.* 2006). The two hominins for whom this method is available, AL 288-1 and KNMWT 15000, are used as a control for the estimation techniques.

When a closer look is taken at the estimates for AL 288-1, a similar picture as presented in Geissmann (1986a) arises. Many of the "standard human"-based equations overestimate the stature greatly. However, the Pygmy-based equations almost all give a result within 10 cm of the actual result as determined by anatomical reconstruction of (Schmid 1983, in Geissmann (1986a)). The ratio method gives a spot-on result, while some "standard human" and nonhuman hominoid-based equations come close too. This is a strong signal that indeed extrapolation increases the likelihood of false results, and the choice of a similarly-sized reference sample aids in increasing reliability.

The publishing authors of KNM-WT 15000 chose not to reconstruct its stature with the anatomical method, but to use regression equations (Ruff and Walker 1993). Ohman *et al.* (2002) did use the anatomical method to reconstruct its stature and came to lower results ("*conservatively (..) at about 147 cm*", Ohman *et al.* 2002, 134) than all other methods, both regression and ratio, that have been used here to calculate the stature of KNM-WT 15000. This could only be possible if KNM-WT 15000 did have very deviant body proportions (a short vertebral column to femur), as proposed by Ohman *et al.* (2002, 131) due to either its peripubescent phase or an abnormality. This shows the complete dependency on shared similarities between goal individual and reference sample for the accuracy of the estimation. However, new material has come available for KNM-WT 15000 (Haeusler *et al.* 2011), restoring somewhat of the perceived abnormality (e.g.the 6th lumbar vertebra) to the normal situation as it is in modern humans. This might (partially) amend the picture of the body proportions painted by Ohman *et al.* (2002), thereby also possibly influencing their deviant body size estimates.

The stature estimation of one of the Dmanisi femora (D4167) by Lordkipanidze *et al.* (2007) was assessed. It was estimated, as adjusted by Ruff (2010), as 144.3–155.7 cm (without confidence intervals). The publishers used the equations from Trotter and Gleser (1952) and from Lundy and Feldesman (1987). As the femur's size is with 38 cm very close to the average of the Pygmy sample while it is below the modern American's average (as used by Trotter and Gleser (1952), the Pygmy equations from the dataset were used to calculate the stature of D4167. The results of this gave the indication that perhaps the stature of this individual was more probable to be between 143–152 cm. The ratio equation agrees with the lower part of this estimate at 144.4 cm. This result would still need to have confidence intervals applied.

The results from these three approaches should give more insight in the stature estimation dynamics of hominins. However, main caveats apply. Although the total amount of gathered equations was well over a hundred, a heavy selection was necessary to reduce over-representation of samples and filtering out possible noise. The actual dataset on which tests could be executed therefore became much smaller. This resulted in a total amount of samples per group as low as $N = 2$. Another problem is the availability of reference samples from some parts of the world. The research intensity has varied between different areas. Europe (Bach 1965; Boldsen 1984; Breitinger 1937; Eliakis *et al.* 1966; Hauser *et al.* 2005; Lorke *et al.* 1953; Maijanen and Niskanen 2010; Mollison 1911; Olivier 1963; Olivier *et al.* 1978; Olivier and Tissier 1975; Pearson 1899; Rother 1971; Telkkä 1950) and the USA (Dupertuis and Hadden 1951; Hens *et al.* 2000; Ousley 1995; Trotter and Gleser 1952, 1958; Wilson *et al.* 2010) have seen a large share of the attention, while information on African (Feldesman and Lundy 1988; Raxter *et al.* 2008), Pygmy African (Hens *et al.* 2000; Jungers 1988; Olivier 1976), Asian (Stevenson 1929), indigenous American (Béguelin 2009; Genoves 1967; Ross and Manneschi 2011; Sciulli and Giesen 1993) and Australian populations has been much more scarce. Such data problems seriously hamper the trust that can be placed in the results.

8.2 Stature estimation of *Homo erectus*

Using the results of this and other research, what can be said about the method of preference for stature estimation in *Homo erectus*?

Any estimation equation is based on a population, the reference sample. Using such an equation on an individual not belonging to that population means estimating the stature of that individual as if it was part of the reference sample. If a hominin with a femur of 45 cm had body proportions like a modern human with a like-sized femur, stature would be reasonably accurately estimated when using an equation based on this modern human population.

The probable error of *Homo erectus* estimates is theorized to be lower when the reference sample is more similar to the goal individual as possible, on the level of total body size, body proportions, functioning and "lifestyle". However, the true reliability of estimation methods for hominins can hardly be tested. The best situation for a control would be an adult skeleton with all bones that contribute to stature present and complete. This would reduce the uncertainties to the minimum that is possible, although assumptions on the allowances that need to be made for soft tissue and spinal curvatures will still be necessary and the validity of extending the conclusions based on that one individual to all others of the species remains to be seen.

In most cases however, a single fossil femur is found, leaving no options for a control in the form of more bones from the same skeleton. For these fossils, a stature estimate will always have an unknown amount of uncertainty and error attached to it. However good and appropriate a stature estimate seems, it must always be remembered that no currently living population is an exact copy of an extinct species and there are few opportunities to test assumptions on body proportions or other size and shape characteristics (Konigsberg *et al.* 1998). Furthermore, the rate of fragmentation and distortion of a fossil can make an accurate femur length measurement a challenge, thereby introducing extra potential error into the stature estimation based on that femur length estimate.

It can be said that none of the modern human populations are a mechanically appropriate reference sample for hominins before 50 kya, as skeletal robustness has declined world-wide since this time (Ruff and Walker 1993, 255). Furthermore, there are few reference samples which suffice to all quality criteria (see Chapter 4), especially body mass reference samples (Konigsberg *et al.* 1998; Porter 2002).

When the three main methods of stature estimation (anatomical reconstruction, regression and ratio) are compared, the anatomical method comes out as the most reliable method (Lundy 1985; Maijanen 2009; Raxter *et al.* 2006), but as unfortunately hardly applicable in the fossil record due to scarceness of (semi-) complete skeletons. The ratio method is less well founded mathematically than regression methods (Hens *et al.* 2000), but it has been shown to give good results even for AL 288-1, whose femur is much smaller than any modern human femur, and it does not require a choice between reference populations as it is based on a world-wide sample (Feldesman and Fountain 1996; Feldesman *et al.* 1990). Regression methods always require a reference sample, and therefore always require a choice but the flip side of this is that they are better fitted to their reference sample (Hens *et al.* 2000).

The results from this research give an indication that the reference sample has the largest influence on the estimate. This is partly due to differences in absolute stature between the populations, and in part probably due to some genetic factors which might be related to climatic adaptations (Ruff 2002), to other adaptations (Kurki *et al.* 2008) or simply to relatedness. As the minimizing of differences between the reference sample and the goal individual reduces the likely error, these results would reason for concentrating on the reference sample as most influential factor. To avoid the effects of extrapolation, it can be recommended to use a sample of which the femur length average is close to that of the fossil femur. More research is needed to filter out the exact effects of climate, environment and other factors in different regions (Kurki *et al.* 2008) and the likelihood that *Homo erectus* was affected by these factors as well (Simpson *et al.* 2008).

Both the ratio and regression methods are based on the similarity of body proportions between reference sample and goal individual (Béguelin 2009, 2). It is important to remember this when the stature of a hominin is estimated, or when the estimate is used in further research. The modern humans that are used as reference sample do not belong to the same species as the *Homo erectus* individual, and the body proportions of early hominins are known only from very few and incomplete skeletons. The result of an estimation equation will be an exact number, which gives the illusion of precision and accuracy. However, even for estimation in the reference sample itself a confidence interval applies. This is even more so the case when the equations are applied to a fossil hominin. The error is caused by individual factors such as environmental influences during ontogeny, also at play in modern humans (Hauser *et al.* 2005), the unknown adaptations of the population (e.g., climatic Ruff (1994), energetic Kurki *et al.* (2008)) and the unknown amount of error arising from the application of an equation based on one species to another (Smith 1996). The size of the confidence interval however cannot really be calculated, as this would require a sample of *Homo erectus* individuals with known stature. At the very least the confidence interval that is applicable for the reference sample itself need to be used. As error ranges are increased around the estimate, the question about the biological meaningfulness of the stature estimate rises. When a confidence interval gets too wide, the estimates will be useless in further calculations that use the estimated stature (e.g. for energetics, Aiello and Key 2002; Aiello and Wells 2002). These problems will need to be assessed by all who use body size estimations for further investigations. They will be aided by new finds of *Homo erectus* postcranial material, which broaden the knowledge on the variation in size and proportions within *Homo erectus* thus helping to minimize error by scraping off a little of the “unknown” and aiding the choice of a reliable estimation method. Body size estimates of hominins however, will always have to be treated as indications.

Appendix A

Appendix

Abstract

The estimated body sizes of hominin individuals and the averages per species are used in palaeoanthropological research to gain more insight in the biology, ecology and life history of *Homo erectus* (Aiello and Key 2002; Foley 1987; McHenry and Coffing 2000). For such inferences to be made, body size needs to be estimated reliably and accurately. The methods that are used in body stature estimation by the use of femur length are evaluated in this thesis.

For this evaluation, 110 stature estimation equations for a range of modern human populations were collected. These equations were used for the calculation of stature estimates for 25 hominin fossils with a range of femur lengths. These estimates are presented in the appendix. The reliability of techniques used for stature estimation on the bases of femur length are evaluated using three different approaches. First, the reference samples that are used for stature estimation in modern humans were evaluated for several quality criteria (Chapter 4). None are flawless, but differences in quality are present. Some of the samples that did not suffice to some of the most important criteria might still be very useful in hominin stature estimation because of their special characteristics, like the Pygmies.

The second approach that was taken was the evaluation of the factors of influence on the formation of the estimation equation (Chapter 5). The mathematical technique that was used to generate the regression was shown to be of minor influence on the regression, while the reference sample did give significant difference between the groups of several geographic/ethnic classifications. This however seemed to be in part related to the inclusion of the Pygmy samples in the comparison. When different stature groups were compared, it was again the smallest group that gave the most deviant results. A conclusion drawn from this is that absolute stature of the reference sample does have an influence on the estimate, and extrapolation out of the size range of the sample indeed increases the likelihood of error (Aiello 1992a; Hens *et al.* 2000; Olivier 1976) as the Pygmy sample of Hens *et al.* (2000) is shown to have a femur/stature ratio that is comparable to other humans and close to the generic human femur/stature ratio (Feldesman *et al.* 1990).

Otherwise, clearly defined groups of populations (African and Caucasian Americans) did

yield significant differences too. Absolute stature therefore is not the only influence on the results. It however was not possible what exactly caused the other differences, be it general relatednes, climatic thermoregulatory adaptation or some other adaptation (Kurki *et al.* 2008; Ruff 2002, 1994). However, caveats apply. The dataset that was used here reflected the differences in past research intensity between regions of the world. Europe, for example, was well represented while other continents (Africa, Asia, Australia) were badly represented amongst the estimation equations. This makes it difficult to make a good comparison. The low number of equations per group also decrease the confidence that can be placed in these results.

In the third approach (Chapter 6), the recommandation following from the second approach was applied to the Dmanisi femur D4167, and reference samples with average femur lengths that were close to the size of D4167 were used for the estimation of its stature. These estimates indicated that perhaps the stature of this individual was more likely to be in the lower range (± 144–150 cm) of the estimated stature in the original publication (± 145–155 cm, excluding confidence intervals)(Lordkipanidze *et al.* 2007). Such a smaller size was also indicated by the use of the femur/stature ratio, which has been shown to be of very good predictive power in AL 288-1 (Chapter 6, Geissmann 1986a; McHenry 1991a). However, the confidence intervals that will need to be added to these estimates will encompass both ranges and extend beyond them; the confidence intervals that Lordkipanidze *et al.* (2007) supply are on average 20 cm.

A reason for the necessity of large confidence interval is illustrated by the difference between the anatomically reconstructed stature of KNM-WT 15000 by Ohman *et al.* (2002) and the calculated estimates of the estimation equations from the dataset. The anatomically reconstructed stature, in principle the most reliable method for stature estimation with the least error, was shown to be shorter than the results from all equations. This serves as a warning that indeed only if the body proportions of an extinct hominin are similar to that of the reference sample, some faith can be put in the estimate. This principle holds, even if the new material of KNM-WT 15000 will show to have more human-like body proportions (Haeusler *et al.* 2011).

An estimate seems like an exact number, but the real estimate must be a range in order to incorporate the probable error. The error is caused by individual factors such as environmental influences during ontogeny, also at play in modern humans (Hauser *et al.* 2005), the unknown adaptations of the population (e.g., climatic Ruff (1994), energetic Kurki *et al.* (2008)) and the unknown amount of error arising from the application of an equation based on one species to another. Due to the error that is unknown, a confidence interval is hard to calculate but must be wide at the same time. This makes it difficult to use such estimates in a biologically relevant manner, for example for further calculations on a species energetics (Smith 1996). The researchers making such calculations will have to address these issues and the danger of compound error. They will be aided by new finds of

Homo erectus postcranial material (e.g. Simpson *et al.* 2008), which broaden the knowledge on the variation in size and proportions within *Homo erectus* thus helping to minimize error by scraping off a little of the “unknown” and aiding the choice of a reliable estimation method.

Bibliography

Aiello, L., 1992a. Allometry and the analysis of size and shape in human evolution. *Journal of Human Evolution* 22, 127–147.

Aiello, L., 1992b. Body size and energy requirements. In: S. Jones, R. Martin and D. Pilbeam (eds), *The Cambridge Encyclopedia of Human Evolution.* Cambridge: Cambridge University Press, 41–45.

Aiello, L. C. and Key, C., 2002. Energetic consequences of being a *Homo erectus* female. *American Journal of Human Biology* 14, 551–265.

Aiello, L. C. and Wells, J. C., 2002. Energetics and the evolution of the genus Homo. *Annual Review Anthropology* 31, 323–38.

Aiello, L. and Dean, C., 1990. *An introduction to human evolutionary anatomy.* New York: Elsevier Academic Press.

Aiello, L. and Wood, B., 1994. Cranial variables as predictors of hominine body mass. *American Journal of Physical Anthropology* 95, 409–426.

Allbrook, D., 1961. The estimation of stature in British and East African males based on tibial and ulnar bone lengths. *Journal of Forensic Medicine* 8, 15–28.

Antón, S., 2008. Framing the question: Diet and evolution in early *Homo.* In: C. Vinyard, M. J. Ravosa and C. Wall (eds), *Primate Craniofacial Function and Biology.* New York: Springer US, 443–482.

Antón, S. C., 2003. Natural history of *Homo erectus. Yearbook of physical anthropology* 46, 126–170.

Antón, S., Leonard, W. and Robertson, M., 2002. An ecomorphological model of the initial hominid dispersal from Africa. *Journal of Human Evolution* 43, 773–785.

Antón, S., Spoor, F., Fellmann, C. D. and Swisher III, C. C., 2007. Defining *Homo erectus*: Size considered. In: Henke, Rothe and Tattersall (eds), *Handbook of Paleoanthropology.* Vol. 3. Berlin: Springer-Verlag, 1656–1693.

Asfaw, B. and et al, W. H. G., 2002. Remains of *Homo erectus* from Bouri, Middle Awash, Ethiopia. *Nature* 416, 317–320.

Auerbach, B. M. and Ruff, C. B., 2004. Human body mass estimation: A comparison of "morphometric" and "mechanical" methods. *American Journal of Physical Anthropology* 125, 331–342.

Bach, H., 1965. Zur Berechnung der Körperhöhe aus den langen Gliedmassenknochen weiblicher Skelette. *Anthropologischer Anzeiger* 29, 812–821.

Bennett, M. R., Harris, J. W., Richmond, B. G., Braun, D. R., Mbua, E., Kiura, P., Olago, D., Kibunjia, M., Omuombo, C., Behrensmeyer, A. K., Huddart, D. and Gonzalez, S., 2009. Early hominin foot morphology based on 1.5 million year-old footprints from Ileret, Kenya. *Science* 323, 1197.

Béguelin, M., 2009. Stature estimation in a Central Patagonian prehispanic population: development of new models considering specific body proportions. *International Journal of Osteoarchaeology* 21(2), 150–158.

Boldsen, J., 1984. A statistical evaluation of the basis for predicting stature from lengths of long bones in european populations. *American Journal for Physical Anthropology* 65, 305–311.

Breitinger, E., 1937. Zur Berechnung der Körperhöhe aus den langen Gliedmassenknochen. *Anthropologischer Anzeiger* 14, 249–274.

Byers, S., Akoshima, K. and Curran, B., 1989. Determination of adult stature from metatarsal length. *American Journal of Physical Anthropology* 79, 275–279.

Carlson, K. J., Grine, F. E. and Pearson, O. M., 2007. Robusticity and sexual dimorphism in the postcranium of modern hunter-gatherers from australia. *American Journal of Physical Anthropology* 134, 9–23.

Charnov, E. L., 1993. Is maximum sustainable yield independent of body size for mammals (and others)? *Evolutionary Ecology* 7(3), 309–311.

Charnov, E. L. and Berrigan, D., 1993. Why do female primates have such long lifespans and so few babies? or life in the slow lane. *Evolutionary Anthropology* 1(6), 191–194.

Conroy, G. C., 2005. *Reconstructing human origins: a modern synthesis.* 2nd edition. New York/London: W.W. Norton & Company, Inc.

Damuth, J. and MacFadden, B. J., 1990. Introduction: body size and its estimation. In: J. Damuth and B. J. MacFadden (eds), *Body size in mammalian paleobiology - Estimation and biological implications.* Cambridge: Cambridge University Press, 1–10.

Dayal, M. R., Steyn, M. and Kuykendall, K. L., 2008. Stature estimation from bones of South African whites. *South African Journal of Science* 104, 124–128.

de Lumley, H., Nioradzé, M., Barsky, D., Cauche, D., Celiberti, V., Nioradzé, G., Notter, O., Zvania, D. and Lordkipanidze, D., 2005. Les industries lithiques préoldowayennes du début du Pléistocène inférieur du site de Dmanisi en Géorgie. *l'anthropologie* 109, 1–182.

Didia, B. C., Nduka, E. C. and Adele, O., 2009. Stature estimation formulae for nigerians. *Journal of Forensic Sciences* 54(1), 20–21.

Dubois, E., 1894. Pithecantropus erectus*: Eine Menschenaehnliche Vebergangsform aus Java.* Batavia: Landesdruckerei.

Dupertuis and Hadden, 1951. On the reconstruction of stature from long bones. *American Journal for Physical Anthropology* 9.

Duyar, R. and Penil, C., 2003. Body height estimation based on tibia length in different stature groups. *American Journal of Physical Anthropology* 22, 23–27.

Eliakis, D., Eliakis, E. and Jordanidis, P., 1966. Détermination de la taille d'après la mensuration des os longs. *Annales de Médicine Légale* 46, 403–421.

Feldesman, M., 1992. Femur/stature ratio and estimation of stature in children. *American Journal of Physical Anthropology* 87, 447–459.

Feldesman, M., Lundy, J. and Kleckner, J., 1989. The femur-stature ratio and estimates of stature in mid- to late-Pleistocene fossil hominids. *American Journal for Physical Anthropology* 78, 219–220.

Feldesman, M. R. and Fountain, R. L., 1996. "Race" specifity and the femur/stature ratio. *American Journal for Physical Anthropology* 100, 207–224.

Feldesman, M. R., Kleckner, J. G. and Lundy, J. K., 1990. Femur/stature ratio and estimates of stature in mid- and late-Pleistocene fossil hominids. *American Journal of Physical Anthropology* 83, 359–372.

Feldesman, M. R. and Lundy, J. K., 1988. Stature estimates for some African Plio-Pleistocene fossil hominids. *Journal of Human Evolution* 17, 583–596.

Ferring, R., Oms, O., Agusti, J., Berna, F., Nioradze, M., Shelia, T., Tappen, M., Vekua, A., Zhvania, D. and Lordkipanidze, D., 2011. Earliest human occupations at Dmanisi (Georgian Caucasus) dated to 1.85-1.78 Ma . *Proceedings of the National Academy of Sciences* 108(26), 10432–10436.

Field, A., 2009. *Discovering Statistics using SPSS (and sex and drugs and rock 'n' roll).* 3rd edition. London: Sage Publications ltd.

Foley, R., 1987. *Another unique species: patterns in human evolutionary ecology.* Harlow, Essex, England: Longman Group UK Limited.

Formicola, V., 1993. Stature reconstruction from long bones in ancient population samples: an approach to the problem of its reliability. *American Journal for Physical Anthropology* 90, 351–358.

Fortelius, 1990. Problems with using fossil teeth to estimate body sizes of extinct mammals. In: Damuth and Macfadden (eds), *Body Size in Mammalian Paleobiology Estimation and Biological Implications.* Cambridge: Cambridge University Press, 207–228.

Fully, G., 1956. Une nouvelle méthode de la taille. *Annales de Médicine Légale* 26, 266–273.

Fully, G. and Pineau, H., 1960. Détermination de la stature au moyen du squelette. *Annales de Médicine Légale* 40, 145–153.

Geissmann, T., 1986a. Estimation of australopithecine stature from long bones: Al 288-1 as a test case. *Folia Primatologica* 47, 119–127.

Geissmann, T., 1986b. Length estimate for knm-er 736, a hominid femur from the lower pleistocene of east africa. *Human Evolution* 1, 481–493.

Genoves, S., 1967. Proportionality of the long bones and their relation to stature among Mesoamericans. *American Journal for Physical Anthropology* 26, 67–78.

Gilbert, W. H., 2008. Hominid systematics. In: W. H. Gilbert and B. Asfaw (eds), Homo erectus - *Pleistocene evidence from the Middle Awash, Ethiopia.* The Middle Awash Series. Berkeley/ Los Angelos/ London: University of California Press, 349–372.

Graves, R. R., Lupo, A. C., McCarthy, R. C., Wescott, D. J. and Cunningham, D. L., 2010. Just how strapping was KNM-WT 15000? *Journal of Human Evolution* 59(5), 542–554.

Grine, F., Jungers, W., Tobias, P. and Pearson, O., 1995. Fossil *Homo* femur from Berg Aukas, northern Namibia. *American Journal for Physical Anthropology* 97, 151–185.

Haeusler, M., Schiess, R. and Boeni, T., 2011. New vertebral and rib material point to modern bauplan of the Nariokotome *Homo erectus* skeleton. *Journal of Human Evolution* 61(5), 575–582.

Hartwig-Scherer, S., 1993. Body weight prediction in early fossil hominids: towards a taxon-'independent' approach. *American Journal of Physical Anthropology* 92, 17–36.

Harvey, P. H. and Clutton-Brock, T., 1985. Life history variation in primates. *Evolution* 39(3), 559–581.

Hauser, R., Smolinski, J. and Gos, T., 2005. The estimation of stature on the basis of measurements of the femur. *Forensic Science International* 147, 185–90.

Hens, S. M., Konigsberg, L. W. and Jungers, W. L., 2000. Estimating stature in fossil hominids: which regression model and reference sample to use? *Journal of Human Evolution* 38, 767–784.

Hernández, M., Garcia-Moro, C. and Lalueza-Fox, C., 1998. Brief communication: Stature estimation in extinct aónikenk and the myth of patagonian gigantism. *American Journal for Physical Anthropology* 105, 545–551.

Holliday, T. W. and Ruff, C. B., 1997. Ecogeographical patterning and stature prediction in fossil hominids: Comment on M.R. Feldesman and R.L. Fountain. *American Journal of Physical Anthropology* 100, 207–224.

Jantz, R., Hunt, D. R. and Meadows, L., 1994. Maximum length of the tibia: How did Trotter measure it? *American Journal of Physical Anthropology* 93, 525–528.

Johanson, D. C. and Edgar, B., 1996. *From Lucy to language.* New York: Nevraumont Publishing Company.

Jungers, W., 1988. Lucy's length: Stature reconstruction in *Australopithecus afarensis* (A.L.288-1) with implications for other small-bodied hominids. *American Journal of Physical Anthropology* 76, 227–231.

Jungers, W. L., 1990. Problems and methods in reconstructing body size in fossil primates. In: J. Damuth and B. J. MacFadden (eds), *Body Size in Mammalian Paleobiology: Estimation and Biological Implications.* Cambridge: Cambridge University Press, 103–118.

Kappelman, J., 1996. The evolution of body mass and relative brain size in fossil hominids. *Journal of Human Evolution* 30, 243–276.

Kerkhoff, A. J. and Enquist, B. J., 2009. Multiplicative by nature: why logarithmic transformation is necessary in allometry. *Journal of Theoretical Biology* 257, 519–521.

Kieffer, C. L., 2010. Tibia and fibula stature formulae for modern female populations based on digital radiographic measurements. *Journal of Forensic Sciences* 55(3), 695–700.

Konigsberg, L., Hens, S., Jantz, L. and Jungers, W., 1998. Stature estimation and calibration: Bayesian and maximum likelihood perspectives in physical anthropology. *Yearbook of physical anthropology* 41, 65–92.

Kramer, A., 1993. Human taxonomic diversity in the pleistocene: Does homo erectus represent multiple hominid species? *American Journal for Physical Anthropology* 91, 161–171.

Kurki, H., GInter, J., Stock, J. and Pfeiffer, S., 2008. Adult proportionality in small-bodied foragers: A test of ecogeographic expectations. *American Journal of Physical Anthropology* 136, 28–38.

Kurki, H., Ginter, J., Stock, J. and Pfeiffer, S., 2010. Body size estimation of small-bodied humans: Applicability of current methods. *American Journal of Physical Anthropology* 141(2), 169–180.

Leakey and Walker, 1985. *Homo erectus* unearthed. *National Geographic* 168, 624–629.

Lieberman, D. E., Devlin, M. J. and Pearson, O. M., 2001. Articular area responses to mechanical loading: Effects of exercise, age, and skeletal location. *American Journal of Physical Anthropology* 116, 266–277.

Lordkipanidze, D., Jashahvili, T., Vakua, A., de León, M. S. P., Zollikofer, C. R., Rightmire, G. P., Pontzer, H., Ferring, R., Oms, O., Tappen, M., Bukhsianidze, M., Agusti, J., Kahlke, R., Kiladze, G., Martinez-Navarro, B., Mouskhelishvili, A., Nioradze, M. and Rook, L., 2007. Postcranial evidence from early homo from dmanisi, georgia. *Nature* 449, 305–310.

Lorke, D., Münzner, H. and Walter, E., 1953. Zur Rekonstruktion der Körpergrösse eines Menschen aus den langen Gliedmassenknochen. *Deutsche Zeitschrift für gerichtliche Medizin* 42, 189–202.

Lundy, J., 1985. The mathematical versus anatomical methods of stature estimation from long bones. *American Journal of Forensic Medicine and Pathology* 6, 73 – 75.

Lundy, J. and Feldesman, M., 1987. Revised equations for estimating living stature from the long bones of South African negroes. *South African Journal of Science* 83, 54–55.

Maat, G. and Mastwijk, 2005. *Manual for the physical anthropological report.* 6 edition. Barge's anthropologica.

Maijanen, H., 2009. Testing anatomical methods for stature estimation on individuals from the W.M. Bass donated skeletal collection. *Journal of Forensic Science* 54(4), 746–752.

Maijanen, H. and Niskanen, M., 2010. New regression equations for stature estimation for medieval Scandinavians. *International Journal of Osteoarchaeology* 20, 472–480.

Manouvrier, L., 1892. *La Détermination de la taille d'après les grands os des membres.* Paris: G. Masson.

McHenry, H., 1974. How large were the *Austalopithecines. American Journal of Physical Anthropology* 40, 329–340.

McHenry, H., 1991a. Femoral lengths and stature in Plio-Pleistocene hominids. *American Journal of Physical Anthropology* 85, 149–158.

McHenry, H. M., 1991b. Petite bodies of the "robust" austalopithecines. *American Journal of Physical Anthropology* 86, 445–454.

McHenry, H. M., 1992. Body size and proportions in early hominids. *American Journal of Physical Anthropology* 87, 407–431.

McHenry, H. M., 1994. Behavioral ecological implications of early hominid body size. *Journal of Human Evolution* 27, 77–87.

McHenry, H. M. and Coffing, K., 2000. *Australopithecus* to *Homo*: transformations in body and mind. *Annual Review Anthropology* 29, 125–146.

Meadows-Jantz, L. and Jantz, R., 1999. Secular change in long bone length and proportion in the United States, 1800 - 1970. *American Journal for Physical Anthropology* 110, 57–67.

Mollison, T., 1911. Die körperproportionen der primaten. *Gegenbaurs Morphologisches Jahrbüch* 42, 79–304.

Murdoch University, 2009. Notes on boxplots. Website. Accessed June 2nd, 2012.
URL: *http://www.cms.murdoch.edu.au/areas/maths/statsnotes/samplestats/boxplot.html*

Nakatsukasa, M., Pickford, M., Egi, N. and Senut, B., 2007. Femur length, body mass, and stature estimates of *Orrorin rugenensis*, a 6 Ma hominid from Kenya. *Primates* 48, 171–178.

Ohman, J., Wood, C., Wood, B., Crompton, R., Günther, M., Yu, L., Savage, R. and Wang, W., 2002. Stature-at-death of KNM-WT 15000. *Human Evolution* 17, 129–142.

Olivier, G., 1963. L'estimation de la stature par les os longs des membres. *Bulletins et Mémoires de la Société d'anthropologie de Paris* XI(4), 433–449.

Olivier, G., 1976. The stature of australopithecines. *Journal of Human Evolution* 5, 529–534.

Olivier, G., Aaron, C., Fully, G. and Tissier, G., 1978. New estimations of stature and cranial capacity in modern man. *Journal of Human Evolution* 7, 513–518.

Olivier, G. and Tissier, H., 1975. Estimation de la stature féminine d'après les os longs des membres. *Bulletins et Mémoires de la Société d'anthropologie de Paris* XIII(2), 297–305.

Ousley, S., 1995. Should we estimate biological or forensic stature? *Journal of Forensic Sciences* 40(5), 768–773.

Packard, G. C., Birchard, G. F. and Boardman, T. J., 2010. Fitting statistical models in bivariate allometry*Biological Reviews* 000–0000.

Pan, 1924. Length of long bones and their proportion to body height in hindus. *J. Anat.* 58.

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* 192, 101–144.

Pelin, C., Zagyapan, R., Yazici, C. and Kürkcüoglu, A., 2010. Body height estimation from head and face dimensions: a different method. *Journal of Forensic Sciences* 55(5), 1326–1330.

Plavcan, J. M., Kay, R. F., Jungers, W. L. and van Schaik, C. P. , 2002. *Conclusions: Reconstructing Behavior in the Fossil Record.* Kluwer Academic / Plenum Publishers. 413–428.

Porter, A., 1999. The prediction of physique from the skeleton. *International Journal of Osteoarchaeology* 9, 102–115.

Porter, A., 2002. Estimation of body size and physique from hominin skeletal remains. *HOMO* 53, 17–38.

Potts, R., Behrensmeyer, A. K., Deino, A., Ditchfield, P. and Clark, J., 2004. Small Mid-Pleistocene hominin associated with East African Acheulean technology. *Science* 305(75), 75–78.

Radoinova, D., Tenekedjiev, K. and Yordanov, Y., 2002. Stature estimation from long bone lengths in bulgarians. *Homo* 52, 221–32.

Raxter, M. H., Auerbach, B. M. and Ruff, C. B., 2006. Revision of the Fully technique for estimating statures. *American Journal of Physical Anthropology* 130, 374–384.

Raxter, M. H., Ruff, C. B. and Auerbach, B. M., 2007. Technical note: Revised fully stature estimation technique. *American Journal of Physical Anthropology* 133, 817–818.

Raxter, M. H., Ruff, C. B., Azab, A., Erfan, M., Soliman, M. and El-Sawaf, A., 2008. Stature estimation in ancient Egyptians: A new technique based on anatomical reconstruction of stature. *American Journal for Physical Anthropology* 136, 147–155.

Rightmire, G. P., Lordkipanidze, D. and Vekua, A., 2006. Anatomical descriptons, comparative studies and evolutionary significance of the hominin skulls from Dmanisi, Republic of Georgia. *Journal of Human Evolution* 50, 115–141.

Robson, S. L. and Wood, B., 2008. Hominin life history: reconstruction and evolution. *Journal of Anatomy* 212, 394–425.

Rollet, E., 1888. *De la mensuration des os longs des membres dans ses rapports avec anthropologie.* PhD thesis. La Faculté de Médicine et de Pharmacie de Lyon. Lyon.

Ross, A. H. and Manneschi, M. J., 2011. New identification criteria for the Chilean populations: estimation of sex and stature. *Forensic Science International* 204, 206.e1–3.

Ross, A. and Konigsberg, L., 2002. New formulae for estimating stature in the balkans. *Journal of Forensic Science* 47, 165–7.

Rother, B., 1971. *Möglichkeiten und Grenzen der Körperhöhenrekonstruktion aus den Massen langer Röhrenknochen.* PhD thesis. University of Leipzig. Leipzig.

Ruff, C., 1988. Hindlimb articular surface allometry in hominoidea and macaca, with comparisons to diaphyseal scaling. *Journal of Human Evolution* 17, 687–714.

Ruff, C. , 1990. *Body mass and hindlimb bone cross-sectional and articular dimensions in anthropoid primates.* Cambridge University Press. 119–49.

Ruff, C., 1991. Climate and body shape in hominid evolution. *Journal of Human Evolution* 21, 81–105.

Ruff, C., 2002. Variation in human body size and shape. *Annual Review of Anthropology* 31, 211–232.

Ruff, C., 2007. Body size prediction from juvenile skeletal remains. *American Journal of Physical Anthropology* 133, 698–716.

Ruff, C., 2010. Body size and body shape in early hominins - implications of the gona pelvis. *Journal of Human Evolution* 58, 166–178.

Ruff, C. B., 1994. Morphological adaptation to climate in modern and fossil hominids. *Yearbook of physical anthropology* 37, 65–107.

Ruff, C. B., Trinkaus, E. and Holliday, T. W., 1997. Body mass and encephaliszation in Pleistocene *Homo*. *Nature* 387, 173–176.

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

Ruff, C. B., Walker, A. and Trinkaus, E., 1994. Postcranial robusticity in homo. iii: Ontogeny. *American Journal of Physical Anthropology* 93, 35–54.

Ryan, I. and Bidmos, M., 2007. Skeletal height reconstruction from measurements of the skull in indigenous south africans. *Forensic Science International* 167, 16–21.

Schmid, 1983. Eine Rekonstruktion des Skelettes von AL 288-1 und deren Konsequenzen. *Folia Primatologica* 40, 283–306.

Schmidt-Nielsen, K., 1984. *Scaling - Why is animal size so important?* New York: Cambridge University Press.

Sciulli, P. W. and Giesen, M. J., 1993. Brief communication: An update on stature estimation in prehistoric Native Americans of Ohio. *American Journal for Physical Anthropology* 92, 395–399.

Silventoinen, K., Sammalisto, S., Perola, M., Boomsma, D. I., Cornes, B. K., Davis, C., Dunkel, L., de Lange, M., Harris, J. R., Hjelmborg, J. V., Luciano, M., Martin, N. G., Mortensen, J., Nisticò, L., Pedersen, N. L., Skytthe, A., Spector, T. D., Stazi, M. A., Willemsen, G. and Kaprio, J., 2003. Heritability of adult body height: a comparative study of twin cohorts in eight countries. *Twin Research* 6(5), 399–408.

Simpson, S. W., Quade, J., Levin, N. E., Butler, R., Dupont-Nivet, G., Everett, M. and Semaw, S., 2008. A female *Homo erectus* pelvis from Gona, Ethiopia. *Science* 322, 1089–1092.

Smith, R. J., 1993. Logarithmic transformation bias in allometry. *American Journal for Physical Anthropology* 90, 215–228.

Smith, R. J., 1996. Biology and body size in human evolution - Statistical inference misapplied. *Current Anthropology* 37(3), 451–481.

Smith, R. J., 2002. Estimation of body mass in paleontology. *Journal of Human Evolution* 43, 271–287.

Smith, R. J., 2009. Use and misuse of the reduced major axis for line-fitting. *American Journal of Physical Anthropology* 140, 476–486.

Smithsonian Institute, 2012. Terry collection. Website. Accessed June 2nd, 2012.
URL: *http://anthropology.si.edu/cm/terry.htm*

Spocter, M. A. and Manger, P. R., 2007. The use of cranial variables for the estimation of body mass in fossil hominins. *American Journal of Physical Anthropology* 134, 92–105.

Spoor, F., Leakey, M., Gathogo, P., Brown, F., Antón, S., McDougall, I., an F.K. Manthi, C. K. and Leakey, L., 2007. Implications of new early *Homo* fossils from ileret, east of Lake Turkana, Kenya. *Nature* 448, 688–691.

Steudel, K., 1980. New estimates of early hominid body size. *American Journal of Physical Anthropology* 52, 63–70.

Steudel-Numbers, K. L., 2006. Energetics in *Homo erectus* and other early hominins: The consequences of increased lower-limb length. *Journal of Human Evolution* 51, 445–453.

Stevenson, J. C., Mahoney, E. R., Walker, P. L. and Everson, P. M., 2009. Technical note: Prediction of sex based on five skull traits using decision analysis (chaid). *American Journal of Physical Anthropology* 139, 434–441.

Stevenson, P. H., 1929. On racial differences in stature long bone regression formulae, with special reference to stature reconstruction formulae for the Chinese. *Biometrika* 21, 303–321.

Subramanian, S., Özaltin, E. and Finlay, J. E., 2010. Height of nations: a socioeconomic analysis of cohort differences and patterns among women in 54 low- to middle-income countries. *PLoS ONE* 6(4).

Swisher III, C. C., Curtis, G. H., Jacob, T., Getty, A. G., Suprijo, A. and Widiasmoro, 1994. Age of the earliest known hominids in Java, Indonesia. *Science* 274, 1870–1874.

Sylvester, A. D., Kramer, P. A. and Jungers, W. L., 2008. Modern humans are not (quite) isometric. *American Journal of Physical Anthropology* 137, 371–383.

Tattersall, I., 2007. *Homo ergaster* and its contemporaries. Berlin: Springer-Verlag, .

Tattersall, I. and Mowbray, K., 2005. Species and paleoanthropology. *Theory in Biosciences* 123, 371–379.

Telkkä, A., 1950. On the prediction of human stature from the long bones. *Acta Anatomica* 9, 103–117.

Trinkaus, E., 1981. Neandertal limb proportions and cold adaptations. In: C. Stringer (ed.), *Aspects of human evolution.* London: Taylor and Francis, 187–224.

Trotter, M., 1970. Estimation of stature from intact long limb bones. In: T. Stewart (ed.), *Personal Identification in Mass Disasters.* Washington, D.C.: National Museum of Natural History, Smithsonian Institution, 71–84.

Trotter, M. and Gleser, G. C., 1951. Trends in stature of American Whites and Negroes born between 1840 and 1924. *American Journal of Physical Anthropology* 9, 427–440.

Trotter, M. and Gleser, G. C., 1952. Estimation of stature from long bones of American Whites and Negroes. *American Journal of Physical Anthropology* 10, 463–514.

Trotter, M. and Gleser, G. C., 1958. A re-evaluation of estimation of stature based on measurements of stature taken during life and of long bones after death. *American Journal of Physical Anthropology* 16, 79–123.

Weidenreich, F., 1939. On the earliest representatives of modern mankind recovered on the soil of east asia. *Peking National Historical Bulletin* 13, 161–174.

Wilson, R. J., Herrmann, N. P. and Jantz, L. M., 2010. Evaluation of stature estimation from the database for forensic anthropology. *Journal of Forensic Science* 55, 684–689.

Wood, B., 2010. Reconstructing human evolution: Achievements, challenges, and opportunities. *Proceedings of the National Academy of Sciences* 107, 8902–8909.

Wood, B., Aiello, L., Wood, C. and Key, C., 1998. A technique for establishing the identity of 'isolated' fossil hominin limb bones. *Jounal of Anatomy* 193, 61–72.

Wood, B. and Collard, M., 1999. The changing face of genus *Homo*. *Evolutionary Anthropology* 8(6), 195–207.

Wurm, H. and Leimeister, H., 1986. Zur geschichte der köperhöhenschätzung nach skelettfunden, teil 3: Ein beitrag zur spezifischen auswahl von vorschlägen zur körperhöhenschätzung nach skelettfunden, zur vergleichbarkeit von schätzergebnissen und zur allgemeinen problematik realistischer lebendhöhenschätzungen. *Anthropologischer Anzeiger/ Gegenbauers morphologischem Jahrbuch* 44/132, 69–110.

List of Figures

List of Tables