CrossMark

ORIGINAL ARTICLE

Sex estimation from the long bones of modern South Africans

Gabriele C. Krüger[1] · Ericka N. L'Abbé[1] · Kyra E. Stull[2]

Abstract Best scientific practice for sex estimation incorporates accurate techniques that employ appropriate standards and population- and period-specific data. Single measurements provide accurate sex estimations, but multiple measurements and multivariate techniques offer greater validity to biological profile assessments. Appropriate, modern standards for sex estimation are limited to the cranium in South Africans (SA), which warrants the examination of the potential for sex estimation using the postcrania of socially defined SA blacks, whites and coloureds through multivariate models and advanced statistical techniques. A total of 39 standard osteometric measurements were taken from the postcrania of 360 socially defined SA blacks, whites and coloureds (equal sex and ancestry). Univariate and multivariate models were evaluated. Multivariate models, with cross-validation and equal priors, were explored with linear and flexible discriminant analysis (LDA and FDA, respectively). Classification accuracies associated with univariate models ranged from 56 to 89%, whereas multivariate classification accuracies using bone models (i.e. all measurements from one element) ranged from 75 to 91%. The highest correct classifications were achieved with multivariate subsets (i.e. combinations of measurements from different bones) and ranged from 90 to 98%. Overall, FDA and LDA yielded similar accuracy rates. Postcranial bones achieve comparable classification accuracies to the pelvis and higher accuracies than metric or morphological techniques using the cranium. While LDA is the most commonly used classification statistic in biological anthropology, FDA provides a good alternative for classification.

Keywords Flexible discriminant analysis · Sexual dimorphism · Postcranial · Advanced statistical analyses · Sex classification

Introduction

Best practice guidelines for sex estimation, as set out by the Scientific Working Group for Anthropology (SWGANTH), advocate that the most reliable and objective morphoscopic and osteometric techniques be applied to an unknown case. Applicable standards and population-specific and period-specific data should also be employed [1, 2]. Postcraniometric methods to estimate sex in US populations show promising results with univariate and multivariate bone models providing classification accuracies up to 95% [2, 3]. Past studies on the postcrania of South African populations also support the use of univariate sectioning points to estimate sex and suggest the method is useful in cases of fragmentary remains [4–6]. Sex estimation is possible using a univariate approach, but the method is less precise than a multivariate approach because of the limited amount of human variation expressed in a single skeletal measurement. A problem with previously published multivariate discriminant functions is that they require a specific set of measurements in order to run the analysis. If any of the measurements are unavailable, the discriminant function cannot be used and a sex estimate cannot be made. Therefore, the use of a suitable software programme is required for a more accurate sex estimation, particularly when remains are incomplete or fragmented.

✉ Gabriele C. Krüger
gabi.kruger@up.ac.za

[1] Department of Anatomy, Faculty of Health Sciences, University of Pretoria, Private Bag X323, Arcadia 0007, South Africa

[2] Department of Anthropology, University of Nevada, 1664 North Virginia Street, Reno, NV 89557-0096, USA

Springer

Reference samples are also limited to black and white South Africans and do not assess the South African coloured group, resulting in an inapplicable standard for this population. The term 'coloured South African' refers to a self-identified group of people who emerged as a result of a complex history of slavery and inter-racial marriages in the country [7]. Coloured South Africans are the second largest population in South Africa (8.9%), but represent the largest group in the Northern Cape and Western Cape, where they make up between 40.3 and 48.8 % of the population, respectively [8, 9].

The Daubert criteria for 'good science' mandates biological anthropology researchers to investigate scientifically tested methods with potential error rates [10, 11]. For sex estimation, many osteometric and morphologic techniques have been updated with more advanced statistical analyses, as well as probabilities and accuracies, for most of the skeletal elements [10, 12–16]. Software programmes that perform multivariate statistical analyses, such as Fordisc3.1 (FD3.1), CRANID and OSSA, are regularly used in forensic anthropology laboratories around the world. A limiting factor of discriminant function analysis, which is a common statistic employed in anthropological software programmes, is the statistical assumption that the unknown belongs to one of the reference populations in the programme. Currently, not all populations are represented in the Forensic Databank [17]. A database was created from cranial data of white, black and coloured South Africans for use in FD3.1 to ensure the same statistical rigour could be applied [16]. However, the postcranial skeleton of South Africans needs to be evaluated with regard to their accuracies in estimating sex and ancestry.

Linear discriminant analysis is a consistent and reliable classification technique. Yet, new statistical analyses are continually being investigated for use in improving classifications. Flexible discriminant analysis (FDA) is not commonly used in biological anthropology, but may provide a novel approach to classifying unknown individuals for the creation of estimations of biological parameters. The purpose of this paper is to evaluate the potential of postcranial measurements from the three largest socially defined South African groups to estimate sex.

Materials and methods

The stratified, random sample consisted of 360 peer-reported black, white and coloured South African males and females, with equal numbers in each sex and ancestry stratum. All individuals were between the ages of 18 and 87 years with a mean of 53 years (Table 1). The samples are housed at the Pretoria Bone Collection at the University of Pretoria (Department of Anatomy) and the Kirsten Collection at the University of Stellenbosch (Division of Anatomy and Histology). The skeletal collections are cadaver-based and

consist of unclaimed, but known, and donated individuals for medical training and/or research. Both institutions have actively accessioned skeletons since the 1950s [18, 19]. Data were collected from the three largest socially identified groups present in South Africa to ensure as much population variation is captured in the reference population.

Coloured South Africans are a self-identified group from the Western Cape. Their genetic heritage is composed of, on average, equal contributions from Bantu-speakers, Europeans, Khoesan (indigenous South Africans) and Asians, making them one of the most genetically admixed groups in the world [7, 20–22]. However, genetic contributions vary between the sexes, at the individual level and in geographic location within South Africa [7, 22, 23]. White South Africans are descended largely from colonial immigrants including Dutch, French, British and Germans [24, 25], with recent genetic research demonstrating almost equal contributions from all Europeans [26]. Black South Africans primarily arose from Bantu-speaking groups that migrated from the Nigerian/Cameroon highlands into South Africa within the past 3000 to 5000 years [27–30]. Genetic and morphological differences do exist between the historical Bantu-speaking and indigenous South Africans (Khoesan), despite evidence for some gene flow among certain groups [13, 31, 32].

A total of 39 standard measurements were taken from 11 postcranial bones. Table 2 refers to full measurement names and abbreviations. Measurements were taken from the left side, unless trauma or pathology was present or the bone was missing, in which case the right side was substituted. While an attempt was made to obtain full numbers for each measurement, some data had to be removed post hoc as outliers. All measurements were taken according to the measurement guidelines presented in Data Collection Procedures for Forensic Skeletal Material [33] and abbreviations were taken from FD 3.1 [34]. All statistical analyses were executed in R or FD3.1 [34, 35].

Statistical analyses

The final measurement numbers, the means and the standard deviations for all measurements and each sex-ancestry group are presented in Table 3. The postcranial bones of five individuals were re-measured by the principal investigator and by a second observer to analyse the repeatability of the measures. The inter- and intra-observer error was assessed using absolute technical error of measurement (TEM) and relative technical error of measurement (%TEM) and was visualized using Bland-Altman plots. A high TEM is associated with a large mean and a low TEM is associated with a small mean. The association between TEM and size of measurement is problematic, as comparative imprecision of different measurements cannot be assessed [36]. In order to compare the TEM collected

Table 1 Sample sizes and abbreviations for each ancestry and sex group for modern South Africans

Collection		n	Mean age	Abbreviation
Pretoria bone collection	Black females	60	47.0	BF
	Black males	60	47.8	BM
	White females	60	67.3	WF
	White males	60	61.9	WM
Kirsten Collection	Coloured females	60	44.1	CF
	Coloured males	60	52.6	CM

from different variables, the absolute TEM is converted to relative TEM (%TEM). The relative TEM is calculated in order to obtain the error expressed as percentage corresponding to the size of the measurement [37]. Each Bland-Altman plot illustrates measurement differences either between measurements taken by two separate observers (inter-observer error Bland-Altman plot) or between two rounds of measurements performed by the same observer (intra-observer error Bland-Altman plot) for all 39 variables. The measurement differences between observers (y-axis) are plotted against the measurement means (x-axis) and additional reference lines include the interval of two standard deviations of the measurement differences on either side of the mean difference [38, 39].

Student's t tests were used to compare the male and female measurements when the populations were pooled. A Bonferroni correction was run to counteract any type I errors due to multiple comparisons. Additionally, analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) were used to test the effects of sex, ancestry

and an interaction between sex and ancestry for every individual measurement and for the multivariate subsets, respectively. If a measurement demonstrated a significant difference between the sexes with the Student's t test or the ANOVA results, univariate models were created and assessed.

Linear discriminant analysis (LDA) is the most commonly used type of parametric classification statistic in biological anthropology. Generally, LDA is employed when there is a categorical response variable (i.e. coloured males, white females, etc.) and continuous predictor variables (i.e. measurements). Assumptions for LDA include independent observations, a normal distribution, equal variance covariance matrices, and homoscedasticity. In LDA, a factor or weight is calculated for each measurement, which, when added together for all variables, maximizes the mean differences among groups [40]. Both forward and backward stepwise selection was used to choose only the most discriminatory variables from each subset. The results provide a probability that the unknown belongs to one of the groups in the reference sample

Table 2 Full measurement names and abbreviations

Clavicle maximum length	claxln	Breadth of S1	sacs1b
Clavicle vertical diameter	clavrd	Innominate height	innoht
Clavicle anterior-posterior diameter	claapd	Iliac breadth	iliabr
Scapula height	scapht	Femur maximum length	femxln
Scapula breadth	scapbr	Femur bicondylar length	fembln
Humerus maximum length	humxln	Femur epicondylar breadth	femebr
Humerus epicondylar breadth	humebr	Femur vertical head diameter	femhdd
Humerus vertical head diameter	humhdd	Femur subtrochanteric anterior-posterior diameter	femsap
Humerus maximum midshaft diameter	hummxd	Femur subtrochanteric transverse diameter	femstv
Humerus minimum midshaft diameter	hummwd	Femur midshaft anterior-posterior diameter	femmap
Radius maximum length	radxln	Femur midshaft transverse diameter	femmtv
Radius anterior-posterior diameter	radapd	Tibia condylo-malleolar length	tibxln
Radius transverse diameter	radtvd	Tibia proximal epiphyseal breadth	tibpeb
Ulna maximum length	ulnxln	Tibia distal epiphyseal breadth	tibdeb
Ulna dorso-volar diameter	ulndvd	Tibia maximum diameter at nutrient foramen	tibnfx
Ulna transverse diameter	ulntvd	Tibia transverse diameter at nutrient foramen	tibnft
Ulna physiological length	ulnphl	Fibula maximum length	fibxln
Sacral height	sacaht	Fibula maximum midshaft diameter	fibmdm
Sacral breadth	sacabr	Calcaneus breadth	calcbr

Table 3 Means and standard deviations (SD) for each sex and ancestry group and all measurements

	BF			BM			WF			WM			CF			CM		
	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD	n	Mean	SD
claxln	35	141	8	37	155	8	42	144	8	41	159	10	52	136	7	52	151	10
claapd	35	9	1	39	11	1	46	9	1	41	11	1	55	9	1	55	11	1
clavrd	33	11	1	37	13	1	46	11	1	42	13	2	55	11	1	55	13	1
scapht	50	135	8	54	154	9	55	145	9	50	163	10	54	135	6	44	151	10
scapbr	58	92	6	56	105	7	59	97	6	57	108	6	56	90	4	55	102	6
humxln	60	297	15	60	323	18	58	313	15	59	338	17	58	289	15	58	315	17
humebr	60	56	3	60	62	3	60	57	3	60	65	4	58	53	3	59	61	4
humhdd	60	38	2	59	44	2	59	43	2	58	48	3	57	38	2	60	44	3
hummxd	60	20	1	59	23	2	60	21	2	60	24	2	57	19	2	59	22	2
hummwd	60	15	1	60	18	2	59	16	2	59	19	2	58	15	1	60	18	2
radxln	60	231	14	60	257	15	57	230	12	57	254	12	60	218	12	56	243	14
radapd	60	11	1	60	13	1	57	11	1	57	13	1	60	10	1	59	13	1
radtvd	60	13	1	60	15	1	58	14	2	58	17	2	60	13	1	58	15	1
ulnxln	58	250	14	60	276	16	59	247	13	60	272	13	57	235	13	53	260	15
ulndvd	59	13	1	60	16	2	60	14	2	56	16	2	60	13	1	59	16	2
ulntvd	60	13	1	60	16	2	59	14	1	60	17	2	60	13	1	58	16	1
ulnphl	58	218	13	59	243	14	59	216	11	60	236	12	59	206	12	57	227	15
sacaht	33	97	10	47	104	10	27	105	10	31	108	12	32	97	8	25	99	7
sacabr	46	92	6	54	89	6	49	101	6	54	98	7	38	95	7	44	90	5
sacs1b	40	44	5	53	48	3	51	45	3	45	48	3	39	41	3	29	45	3
innoht	58	189	9	58	205	10	59	209	10	59	224	12	50	185	8	55	201	12
iliabr	56	142	7	58	150	7	59	163	9	54	165	9	51	141	7	58	146	9
femxln	59	425	19	60	457	22	57	441	20	60	471	23	59	413	19	57	446	24
fembln	59	421	19	60	454	22	57	438	21	60	468	23	59	409	19	57	443	24
femebr	60	72	4	60	80	4	57	76	3	58	84	4	58	70	3	57	78	4
femhdd	60	40	2	59	46	2	58	43	2	60	48	3	59	40	2	58	46	3
femsap	59	25	2	57	28	2	52	27	2	52	30	2	55	25	2	56	27	2
femstv	60	28	2	59	32	2	55	31	2	55	33	2	58	29	2	57	33	2
femmap	59	27	2	59	30	2	58	28	2	52	31	2	53	26	2	47	29	2
femmtv	60	25	2	59	27	2	55	26	2	57	29	2	59	24	2	57	27	2
tibxln	56	361	19	58	390	21	57	365	21	59	391	21	59	344	20	56	373	25
tibpeb	58	68	4	58	76	4	55	71	4	57	79	4	57	66	3	52	74	4
tibdeb	58	44	2	58	49	3	58	47	3	59	52	3	57	43	2	55	47	3
tibnfx	57	32	2	53	36	3	56	33	2	55	37	3	55	31	2	50	35	2
tibnft	59	23	3	60	26	2	59	24	2	56	27	2	59	22	2	54	25	2
fibxln	58	353	19	57	383	21	48	358	21	56	383	20	56	336	17	49	365	22
fibmdm	59	14	2	56	16	2	54	15	1	56	16	2	55	14	1	47	15	2
calcxl	52	74	4	58	83	4	46	80	4	48	86	5	0	–	–	0	–	–
calcbr	51	39	2	59	43	2	55	40	2	55	44	3	50	41	3	51	43	4

(posterior probability) [41]. Equal prior probabilities were used and results were cross-validated using leave-one-out cross validation (LOOCV). The LDA function in the MASS package was used to conduct LDA and the stepclass function in the klaR package was used to conduct stepwise variable selection in R [42].

FDA is similar to LDA, except it replaces linear regression with a non-parametric regression method [43]. For this study, FDA built regression models defined by multivariate adaptive regression splines (MARS) [44]. MARS uses piece-wise linear segments to model non-linear relationships. To identify the optimal model, a forward and backward stepwise procedure is

implemented along with automatic variable selection [45]. FDA is considered advantageous to LDA because FDA generates a flexible surface to separate the classes and is not as restricted by assumptions. Subsequently, accuracies have been higher when there are more than two classification categories or when assumptions of normality and homoscedasticity are violated [43]. The FDA function in the mda package was used to conduct FDA in R [46].

Both LDA and FDA were used to classify the sample using the univariate models, bone models, and multivariate subsets. The bone models consist of combinations of the different measurements taken from each bone (Table 4) and the multivariate subsets consist of different combinations of measurements (Table 5). For the purpose of showing the potential of the multivariate approach, only a limited number of multivariate subsets are presented. The selected variables for each model and subset differ slightly between LDA and FDA as seen in Tables 4 and 5. The two multivariate techniques were used to classify the sample according to sex and ancestry (six groups). Only multivariate subsets were utilized and again stepwise selection reduced the number of parameters for each model, as to include only the best measurements for classifying the sample. While some data were missing, the sex and sex-ancestry sample sizes were similar for all groups.

Results

The mean intra-observer TEM and %TEM, for all measurements, were fairly small at 0.13 mm (ranging from 0.0 to 0.82 mm) and 0.24% (ranging from 0.0 to 2.24%), respectively. The mean TEM and %TEM for the inter-observer error were slightly higher at 0.43 mm and 1.04%, and the ranges slightly wider at 0.0 to 0.94 mm and 0.0 to 4.65%, respectively. Measurements associated with the ulna had the greatest error, specifically the dorso-volar and transverse diameters. The Bland-Altman plots of the intra- and inter-observer errors

revealed overall high agreement with no measurement differences exceeding 2 mm or any systematic bias (Fig. 1). Most measurements fell between the upper and lower agreement levels (dashed lines), which were calculated based on the standard deviations. While the means and range of differences did not vary considerably between the intra- and inter-observer errors, the spread of the differences is larger for the inter-observer error than for the intra-observer error.

Student's t tests, with a Bonferroni correction, revealed significant differences ($p < 0.05$) between the sexes for all 39 measurements when ancestry groups were pooled (i.e. all males compared to all females). The ANOVA results demonstrated significant differences ($p < 0.05$) among all three populations for all measurements except claapd, clavrd, ulndvd and calcbr, while the MANOVA results indicated significant differences between sex and ancestry groups for all multivariate subsets. Significant interactions between sex and ancestry were only present for iliabr and calcbr; however, significant interactions were present between sex and ancestry for the upper limb, breadths and all-variable models.

Cross-validated correct classification rates for the univariate models using LDA, when the three ancestry groups were pooled, ranged from 56 to 89%. The radapd and the iliabr classified best and worst, respectively (Table 6). Univariate accuracies using FDA ranged from 60 to 89%, and radapd and sacabr classified best and worst, respectively (Table 6). Cross-validated accuracies for LDA ranged between 75 and 90% for the bone models. The radius performed the best and the sacrum and fibula performed the worst. Females classified better than males for all bone models, except for the ulna and sacrum (Table 7). Cross-validated correct classification accuracies for the bone models ranged from 75 to 91% using FDA. The clavicle was best and the fibula worst at classifying the sample into males and females. Overall, females classified better than males, except for models of the clavicle, humerus, innominate and femur (Table 8).

Cross-validated correct classifications ranged from 90 to 98% when LDA was employed on the multivariate subsets. A sex bias was observed in most subsets with males only achieving higher accuracies for the hip and lower limb bone model subsets (Table 9). Overall classification for the multivariate subsets ranged from 90 to 97% using FDA. The lengths subset achieved the lowest accuracy and the all-variable model achieved the highest accuracy. Females obtained higher correct classifications for all subsets, except for breadths and lengths models (Table 10).

When the sample was classified into six sex-ancestry groups, the number of groups increased considerably, which led to decreased classification accuracies. LDA correctly classified the sample 64 to 80%; the all-variable model achieved the highest classification accuracy, whereas the hip subset achieved the lowest. Overall, white females classified the best and coloured males the worst (Table 11, Fig. 2). A

Table 4 Stepwise selected variables included in bone models for LDA and FDA when classifying according to sex only

	n	LDA	FDA
Clavicle	253	claxln, claapd	claapd, clavrd, claxln
Scapula	294	scapht	scapbr, scapht
Humerus	343	humebr	humebr, hummwd, humhdd
Radius	344	radapd	radapd, radxln
Ulna	339	ulntvd	ulnxln, ulntvd, ulndvd
Sacrum	160	sacabr, sacs1b	sacs1b, sacabr
Innominate	331	innoht, iliabr	innoht, iliabr
Femur	288	femhdd	femhdd
Tibia	293	tibpeb	tibpeb, tibnfx
Fibula	312	fibxln	fibxln, fibmdm

Table 5 Stepwise selected variables included in multivariate subsets for LDA and FDA when classifying according to sex only

	n	LDA	FDA
Hip	130	sacabr, femhdd, femstv	femhdd, sacabr, iliabr, femebr, femmap
Upper limb	183	humhdd, radapd	radapd, claxln, clavrd, claapd, humxln, humhdd
Lower limb	148	sacabr, femhdd, tibpeb	femhdd, sacabr, tibnfx, iliabr, femsap
Breadths	138	scapbr, ulntvd, sacabr	ulntvd, sacabr, radapd, humebr, iliabr, scapbr
Lengths	106	scapht, ulnxln	scapht, radxln, tibxln, claxln, innoht, fibxln
All-variable	124	femhdd, sacabr	ulntvd, femhdd, sacabr, claapd

Fig. 1 Bland-Altman plots representing the intra- (*above*) and inter-observer (*below*) errors, respectively, when both the principal investigator and a second observer re-measured all measurements for the five randomly selected individuals
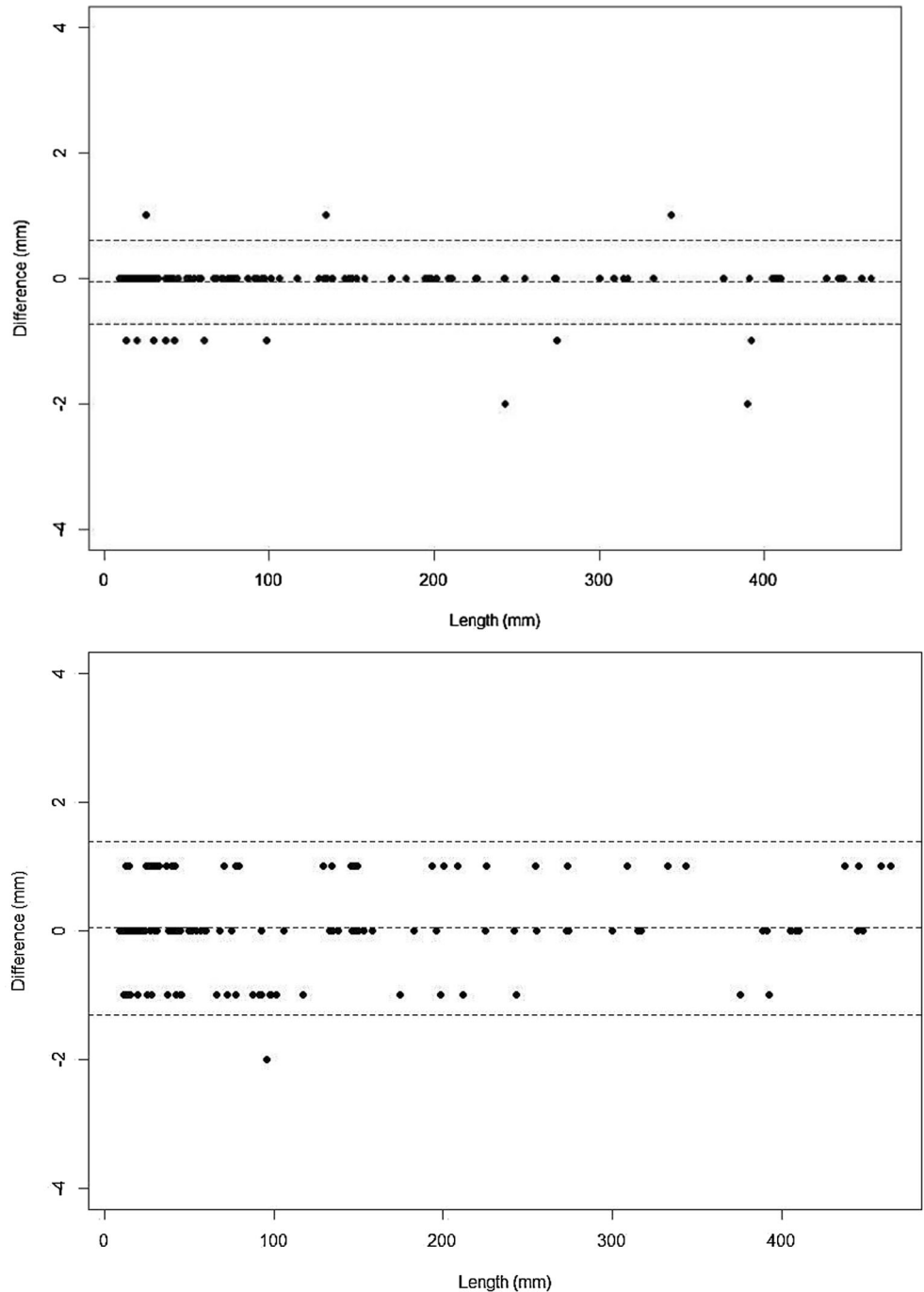
Table 6 Correct classification accuracies (%) for assessment of univariate models using LDA and FDA when ancestry groups are pooled

	LDA	FDA		LDA	FDA
radapd	89	89	fembln	76	76
femhdd	86	86	femmap	76	76
tibpeb	86	85	calcxl	76	76
ulntvd	85	85	clavrd	75	77
humebr	85	84	femxln	75	75
hummwd	83	83	humxln	75	74
scapht	82	82	radtvd	74	74
claapd	81	82	femmtv	74	74
scapbr	81	81	tibxln	73	74
radxln	81	81	femsap	73	73
ulnphl	81	80	calcbr	73	73
femebr	81	79	tibnft	72	76
ulnxln	80	80	tibdeb	70	76
hummxd	79	79	innoht	68	70
ulndvd	79	79	sacs1b	67	72
claxln	78	78	fibmdm	67	67
humhdd	78	78	sacaht	61	61
femstv	78	77	sacabr	60	60
tibnfx	78	75	iliabr	56	64
fibxln	76	77			

Mahalanobis distance matrix for LDA shows the distances between sex-ancestry groups and also shows the distances between males and females of each ancestry (i.e. sexual dimorphism). Coloured South Africans display the greatest amount of sexual dimorphism, while white South Africans display the least amount of sexual dimorphism (Table 12). Slightly lower, but comparable, correct classification accuracies (60–79%) were achieved when FDA classified the sample into sex and ancestry. The lengths subset produced the lowest

Table 7 Classification accuracies for LDA using stepwise selected single bone measurements to estimate sex

	Accuracy (%)			Sex bias (%)
	Males	Females	Combined	
Clavicle	86	90	88	4
Scapula	79	84	82	5
Humerus	84	84	84	0
Radius	89	90	90	1
Ulna	86	85	86	−1
Sacrum	76	74	75	−2
Innominate	73	82	78	9
Femur	85	88	86	3
Tibia	84	88	86	4
Fibula	75	75	75	0

Table 8 Classification accuracies for FDA using bone models to estimate sex

	Accuracy (%)			Sex bias (%)
	Males	Females	Combined	
Clavicle	93	90	91	−3
Scapula	84	87	86	3
Humerus	88	87	88	−1
Radius	89	90	90	1
Ulna	89	89	89	0
Sacrum	74	79	76	5
Innominate	80	77	79	−3
Femur	87	86	87	−1
Tibia	82	90	86	8
Fibula	73	77	75	4

and the all-variable model the highest classification accuracies. Overall, white males classified best and coloured males worst (Table 13).

The FDA plot illustrates the results of a flexible discriminant analysis in discriminant (canonical) coordinates. A function of the method was incorporated to extract discriminant variables, posterior probabilities and to predict class memberships. The observations are classified according to the distance to the nearest centroid in discriminant space by the predict function of the mda package [44, 46]. Interpretation of the relationships between the centroids of all six sex-ancestry groups show overlap between black and coloured females and between black and coloured males, while white males and females are the most distinct of the groups and obtained the highest correct classifications (Figs. 2 and 3). Comparison of the multivariate classification techniques demonstrated that FDA was only slightly better in classifying the sample into sex and ancestry groups with an average accuracy of 70%, compared to the 69% of LDA. The slight disparity could also be explained by sampling variation; however, the lack of large differences between the two methods may also indicate that the data were similar to the assumed distribution.

Table 9 Classification accuracies for LDA using stepwise selected multivariate subsets to estimate sex

	Accuracy (%)			Sex bias (%)
	Males	Females	Combined	
Hip	98	94	96	−4
Upper limb	93	97	95	4
Lower limb	99	95	97	−2
Breadths	98	99	98	1
Lengths	88	92	90	4
All-variable	97	97	97	0

Table 10 Classification accuracies for FDA using multivariate subsets to estimate sex

	Accuracy (%)			Sex bias (%)
	Males	Females	Combined	
Hip	94	96	95	2
Upper limb	95	96	96	1
Lower limb	93	96	95	3
Breadths	98	96	97	−2
Lengths	98	89	93	−5
All-variable	97	97	97	0

Discussion

The type and location of the measurements are important to consider when estimating sex within a population. Of the 39 measurements, taken from each individual in the sample, 26 were breadth measurements. Most variables selected for in model creation in this study were breadth measurements. A multitude of factors contribute to the expression of sexual dimorphism in the skeleton. The current research, as well as the results of numerous publications, suggests that breadth measurements discriminate between the sexes better than length dimensions [47–51].

While the superiority of breadth measurements for the estimation of sex are well documented in biological anthropology, a relatively unexplored area in sex estimation and population studies is whether current measurements and statistical approaches, such as univariate and multivariate models using LDA, are adequately capturing the most sexually dimorphic elements [10, 52, 53]. In using the stepwise selection process on multiple measurements, the likelihood of selecting the measurements with the most discriminating power increases [40]. As in other studies, stepwise selection process and the use of multiple variables provide higher classification accuracies than single measurement models.

Assessment of the bone models and multivariate subsets using both LDA and FDA resulted in similar classification

Table 11 Correct classification rates when sex and ancestry were assessed by LDA using multivariate subsets. Number of variables in subsets is included in parentheses

	n	Accuracy (%)						
		BF	BM	WF	WM	CF	CM	Combined
Hip (4)	228	61	52	92	64	65	57	64
Upper (4)	224	59	64	83	70	68	60	67
Lower (6)	226	49	62	84	73	73	70	68
Lengths (4)	189	57	62	70	81	68	48	65
All-variable (6)	226	78	78	95	81	79	65	80

accuracies for both techniques, with FDA achieving only slightly better results. LDA, as a parametric test, is generally considered more powerful when compared to a non-parametric test, such as FDA. However, LDA is only powerful when assumptions of independence, normality, equal variance covariance matrices and homoscedasticity are met. When assumptions are violated, a non-parametric technique may be more suitable [12, 43]. FDA has resulted in higher overall correct classifications compared to stepwise selected LDA models for assessing sex in juveniles and estimating ancestry from postcranial remains in South Africans [50, 54]. While FDA demonstrated the potential to be a useful classification technique in biological anthropology, no standardized software programme, besides R, exists to use FDA for adult specimens. Because of the extensive number of potential measurement combinations, it is difficult to create discriminant functions for all possible combinations. For a multivariate model, a statistical software programme, such as FD3.1, should be utilized to create the model that yields the highest correct classification based on all available measurements. However, as population-specific databases are required for reliable results to be produced using FD3.1, the data collected for the current study are available for use on South African postcranial remains (available on request from corresponding author).

When the bone model results are compared to the Spradley and Jantz (2011) publication that estimated sex from postcranial remains on a US sample, similar patterns of sexual dimorphism are apparent. In both studies, the clavicle and radius are good sex estimators, whereas the fibula and sacrum are weak estimators. However, overall classification accuracies for South Africans were slightly lower for the bone models when compared to the results obtained by Spradley and Jantz (2011), indicating a lower level of sexual dimorphism among South African groups. While bone models did not prove as useful in the South African context, multivariate subsets far outperformed the bone models of either study.

The all-variable model for classification according to sex and ancestry shows significant sex and ancestry interactions. Thus, ancestry affected the separation of groups. Cranially, South African blacks and coloureds overlap considerably, whereas South African whites tend to be more distinct from the other two groups [7]. A similar outcome is observed in the plot for LDA (Fig. 2) when the all-variable model is used to classify into sex and ancestry. White females had the highest correct classification and the group centroid is located furthest away from the other female centroids (Fig. 2). Misclassifications generally occurred in ancestry, rather than sex, although the white females misclassified by sex, rather than ancestry. Furthermore, the distance between the white female centroid and the black and coloured female centroids is greater than the distances between the group means of the male counterparts. The overall robusticity or size of the white female measurements contributes to the
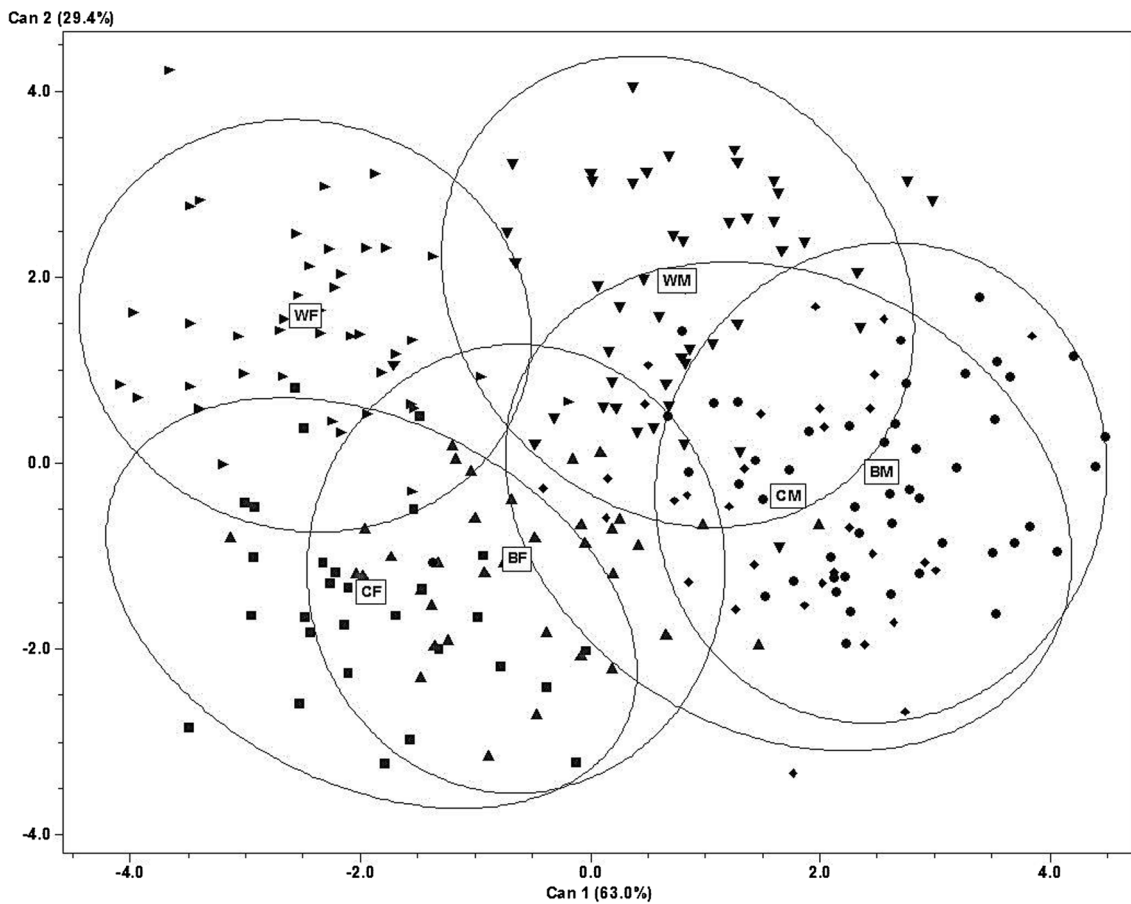
Fig. 2 LDA plot of the all-variable model when classifying according to sex and ancestry (80% correct classification)

separation of the white females from the black and coloured females and the misclassifications of the white females as white males. The mean age of the white females is considerably higher than the other female groups and likely associates with the increase in robusticity [14, 55]. In a previous study involving the morphological expression of sexual dimorphism in the cranium, white females present with a larger number of intermediate/masculine numbers than black females [53]. Furthermore, genetic evidence and population history of white South Africans suggests low rates of intra-group variation for

the whites when compared to black and coloured South Africans and this may have contributed to the distinct separation of white South Africans in comparison to the other South African females.

Conclusion

The current research is a comprehensive postcranial analysis that incorporates a thorough, novel and advanced statistical approach as a means to improve sex estimation standards with

Table 12 Mahalanobis distance[a] matrix for LDA using the all-variable model[b]

	BF	BM	CF	CM	WF	WM
BF	0.0	11.3	4.5	8.6	10.7	12.4
BM	11.3	0.0	22.6	2.3	28.2	7.9
CF	4.5	22.6	0.0	14.7	10.2	18.5
CM	8.6	2.3	14.7	0.0	22.3	6.8
WF	10.7	28.2	10.2	22.3	0.0	11.0
WM	12.4	7.9	18.5	6.8	11.0	0.0

[a] All distances significant at <0.001

[b] All-variable model: ulnxln, ulntvd, sacabr, iliabr, femhdd, and calcbr

Table 13 Correct classification for FDA using multivariate subsets to estimate sex and ancestry

	n	Accuracy (%)						
		BF	BM	WF	WM	CF	CM	Combined
Hip (6)	228	66	58	83	82	66	36	64
Upper limb (4)	224	61	71	67	82	74	71	71
Lower limb (8)	226	71	77	85	82	68	67	75
Lengths (3)	189	49	61	70	80	68	41	60
All-variable (6)	226	79	81	94	85	72	63	79

Number of variables in subsets is included in parentheses

Fig. 3 FDA plot of the all-variable model when classifying according to sex and ancestry (79% correct classification)
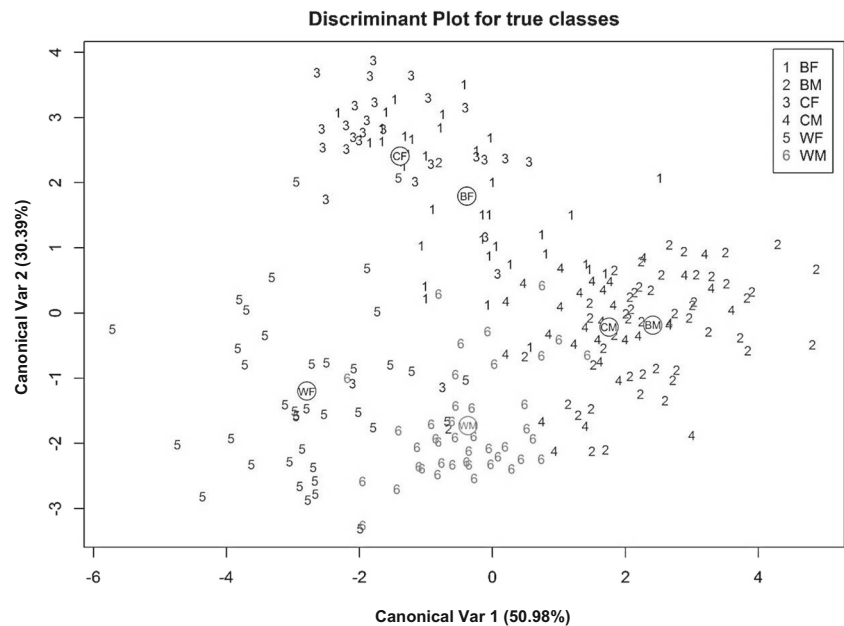
a sample that includes more South African groups, and hence a larger amount of variation than any previous South African sex estimation research. Long bone dimensions are shown to reliably distinguish between males and females in three socially defined South African populations. Postcranial bones achieve comparable classification accuracies to morphological analysis of the pelvis and higher accuracies than metric or morphoscopic techniques using the cranium in South Africa [10, 12, 15, 24, 56, 57]. The current study demonstrates the advantage of multivariate subsets (multiple measurements from multiple bones) for sex estimations from long bones, as the subsets are made up of the most sexually dimorphic elements and thus provide the most reliable results. Furthermore, the resulting database can be used with FD3.1 and applied to forensic cases in South Africa to classify an unknown using the most sexually dimorphic of the available measurements.

References

1. Scientific Working Group for Forensic Anthropology (2010) Sex Assessment. In: Sci. Work. Group Forensic Anthropol. SWGANTH. http://swganth.startlogic.com/. Accessed 27 Jun 2015

2. Spradley MK, Anderson BE, Tise ML (2015) Postcranial sex estimation criteria for Mexican Hispanics. J Forensic Sci 60:S27–S31

3. Spradley MK, Jantz RL (2011) Sex estimation in forensic anthropology: skull versus postcranial elements. J Forensic Sci 56:289–296

4. Steyn M, İşcan MY (1999) Osteometric variation in the humerus: sexual dimorphism in South Africans. Forensic Sci Int 106:77–85

5. Asala SA (2001) Sex determination from the head of the femur of South African whites and blacks. Forensic Sci Int 117:15–22

6. Bidmos M, Asala SA (2003) Discriminant Function sexing of the calcaneus of the South African Whites. J Forensic Sci 48:1–6

7. Stull KE, Kenyhercz MW, L'Abbé EN (2014) Ancestry estimation in South Africa using craniometrics and geometric morphometrics. Forensic Sci Int 245:206.e1–206.e7

8. Statistics South Africa (2011) Census. South Africa. pp 1–88. Available from: www.beta2statssa.gov.za. Accessed 16 August 2014

9. Statistics South Africa (2014) Mid-year population estimates. pp 1–19. Available from: www.beta2statssa.gov.za. Accessed 29 March 2015

10. Klales AR, Ousley SD, Vollner JM (2012) A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. Am J Phys Anthropol 149:104–114

11. Grivas CR, Komar DA (2008) Kumho, Daubert, and the nature of scientific inquiry: implications for forensic anthropology. J Forensic Sci 53:771–776

12. Krüger GC, L'Abbé EN, Stull KE, Kenyhercz MW (2015) Sexual dimorphism in cranial morphology among modern South Africans. Int J Legal Med 129:869–875

13. Liebenberg L, Stull KE, L'Abbé EN, Botha D (2015) Evaluating the accuracy of cranial indices in ancestry estimation among South African groups. J Forensic Sci 60:1277–1282

14. Liebenberg L, L'Abbé EN, Stull KE (2015) Population differences in the postcrania of modern South Africans and the implications for ancestry estimation. Forensic Sci Int 257:522–529

15. Kenyhercz MW (2012) Sex estimation using pubic bone morphology in a modern South African sample: a test of the Klales et. al. method. In: Am. J. Phys. Anthropol. Wiley-Blackwell, pp 179–180

16. L'Abbé EN, Nawrocki SP, Keough N (2013) L'Abbé, Nawrocki, Keough - 2008 - The application of FORDISC 3 . 0 to South African crania.pdf

17. Ramsthaler F, Kreutz K, Verhoff MA (2007) Accuracy of metric sex analysis of skeletal remains using Fordisc® based on a recent skull collection. Int J Legal Med 121:477–482

18. L'Abbé EN, Loots M, Meiring JH (2005) The Pretoria bone collection: a modern South African skeletal sample. HOMO - J Comp Hum Biol 56:197–205

19. L'Abbé EN, Steyn M (2012) The establishment and advancement of forensic anthropology in South Africa. In: Dirkmaat DC (ed) Companion Forensic Anthropol. Wiley-Blackwell, UK, pp 626–638

20. Adhikari M (2005) Contending approaches to coloured identity and the history of the coloured people of South Africa. Hist Compass 3: 1–16

21. Patterson N, Petersen DC, van der Ross RE et al (2010) Genetic structure of a unique admixed population: implications for medical research. Hum Mol Genet 19:411–419

22. Stull KE, L'Abbé EN, Ousley SD (2014) Using multivariate adaptive regression splines to estimate subadult age from diaphyseal dimensions: subadult age estimation from diaphyses. Am J Phys Anthropol 154:376–386

23. Quintana-Murci L, Harmant C, Quach H et al (2010) Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture. Am J Hum Genet 86:611–620

24. Steyn M, İşcan MY (1998) Sexual dimorphism in the crania and mandibles of South African whites. Forensic Sci Int 98:9–16

25. L'Abbé EN, Van Rooyen C, Nawrocki SP, Becker PJ (2011) An evaluation of non-metric cranial traits used to estimate ancestry in a South African sample. Forensic Sci Int 209:195.e1–195.e7

26. Greeff JM (2007) Deconstructing Jaco: genetic heritage of an Afrikaner. Ann Hum Genet 71:674–688

27. Beck RB (2000) The history of South Africa. Greenwood Press, Westport

28. Tishkoff SA, Reed FA, Friedlaender FR et al (2009) The genetic structure and history of Africans and African Americans. Science 324:1035–1044

29. Tishkoff SA, Williams SM (2002) Genetic analysis of African populations: human evolution and complex disease. Nat Rev Genet 3: 611–621

30. May A, Hazelhurst S, Li Y et al (2013) Genetic diversity in black South Africans from Soweto. BMC Genomics 14:644

31. Herbert RK (1990) The sociohistory of clicks in Southern Bantu. Anthropol Linguist 32:295–315

32. Stynder DD (2009) Craniometric evidence for South African Later Stone Age herders and hunter–gatherers being a single biological population. J Archaeol Sci 36:798–806

33. Moore-Jansen PH, Jantz RL (1994) Data collection procedures for forensic skeletal material, 3rd ed. Forensic Anthropology Center, Department of Anthropology, University of Tennessee, Knoxville

34. Jantz RL, Ousley SD (2005) Fordisc 3: computerized forensic discriminant functions. University of Tennessee, Knoxville

35. R Core Team (2013) R: a language and environment for statistical computing. In: R Found. Stat. Comput. Vienna Austria. http://www.R-project.org/

36. Ulijaszek SJ, Kerr DA (1999) Anthropometric measurement error and the assessment of nutritional status. Br J Nutr 82:165–177

37. Perini TA, de Oliveira GL, dos Santos OJ, de Oliveira FP (2005) Technical error of measurement in anthropometry. Rev Bras Med Esporte 11:81–85

38. Fernandez R, Fernandez G (2009) Validating the Bland-Altman method of agreement. In: Annu. Conf. West. Users. SAS Softw, San Jose Calif, pp 1–17

39. Sedgwick P (2013) Limits of agreement (Bland-Altman method). BMJ 346:f1630–f1630

40. Ousley SD, Jantz RL (2012) Fordisc 3 and statistical methods for estimating sex and ancestry. In: Dirkmaat DC (ed) Companion Forensic Anthropol. Wiley-Blackwell, West Sussex, pp 311–329

41. Dawson-Saunders B, Trapp RG (2004) Basic and clinical biostatistics, 4th edn. McGraw-Hill, New York

42. Ripley B (2015) MASS: Support functions and datasets for Venables and Ripley's MASS. http://cran.r-project.org/web/packages/MASS/MASS.pdf. Accessed 29 Jun 2015

43. Hastie T, Tibshirani R, Buja A (1994) Flexible discriminant analysis by optimal scoring. J Am Stat Assoc 89:1255–1270

44. Milborrow S (2013) Notes on the earth package. http://CRAN.R-project.org/package=earth

45. Leathwick JR, Rowe D, Richardson J et al (2005) Using multivariate adaptive regression splines to predict the distributions of New Zealand's freshwater diadromous fish. Freshw Biol 50:2034–2052

46. Hastie T (2013) mda: mixture and flexible discriminant analysis. http://cran.r-project.org/web/packages/mda/index.html

47. Elewa AMT (2010) Morphometrics for nonmorphometricians. Springer Berlin Heidelberg, Berlin Heidelberg

48. Steyn M, İşcan MY (1997) Sex determination from the femur and tibia in South African whites. Forensic Sci Int 90:111–119

49. France DL (1998) Observational and metric analysis of sex in the skeleton. In: Reichs KJ (ed) Forensic Osteol. Adv. Identif. Hum. Remains, vol 2nd. Charles C. Thomas, Springfield, pp 163–186

50. Scheuer L, Black S (2004) The juvenile skeleton. Elsevier Academic Press, New York

51. Stull KE (2014) An osteometric evaluation of age and sex differences in the long bones of South African children from the Western Cape. PhD Dissertation, University of Pretoria

52. Bogin B (1999) Patterns of human growth. In: Camb. Stud. Biol. Anthropol, vol 2nd. Cambridge University Press, Cambridge, p 267

53. Walker PL (2008) Sexing skulls using discriminant function analysis of visually assessed traits. Am J Phys Anthropol 136:39–50

54. Panik MJ (2005) Advanced statistics from an elementary point of view. Academic Press

55. Vance VL, Steyn M, L'Abbé EN (2011) Nonmetric sex determination from the distal and posterior humerus in black and white South Africans. J Forensic Sci 56:710–714

56. L'Abbé EN, Kenyhercz MW, Stull KE et al (2013) Application of FORDISC 3.0 to explore differences among Crania of North American and South African blacks and whites. J Forensic Sci 58:1579–1583

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