

## **Postcraniometric Sex and Ancestry Estimation in South Africa: A Validation Study**

Leandi Liebenberg, MSc<sup>1\*</sup>, Gabriele C. Krüger, MSc<sup>1</sup>, Ericka N. L'Abbé, PhD, D-ABFA<sup>1</sup>, Kyra E. Stull, PhD<sup>1,2</sup>

<sup>1</sup>Department of Anatomy, University of Pretoria, Arcadia, South Africa

<sup>2</sup>Department of Anthropology, University of Nevada, Reno, United States

ORCID:

LL – 0000-0003-4310-5489

GCK – 0000-0002-3289-1122

ENL – 0000-0002-6722-8814

KES - 0000-0002-4541-6777

\* Corresponding author:

Leandi Liebenberg

Department of Anatomy

University of Pretoria

Private Bag x323

Arcadia 0007 South Africa

Tel: +27-82 072 0057/12-420 3256

E-mail: [leandi.liebenberg@up.ac.za](mailto:leandi.liebenberg@up.ac.za)/[leandil@hotmail.co.za](mailto:leandil@hotmail.co.za)

## **Acknowledgements**

The authors wish to thank Mandi Alblas and Linda Greyling from Stellenbosch University for granting access to the Kirsten Collection. Thank you to the reviewers for their helpful comments, which ultimately improved the manuscript. This research was funded by the National Research Foundation (NRF). Any opinions, findings and conclusions or recommendations expressed in the material are those of the authors and therefore the NRF does not accept any liability in regard thereto.

## **Abstract**

With the acceptance of the *Daubert* criteria as the standards for best practice in forensic anthropological research, more emphasis is being placed on the validation of published methods. Methods, both traditional and novel, need to be validated, adjusted, and refined for optimal performance within forensic anthropological analyses. Recently, a custom postcranial database of modern South Africans was created for use in Fordisc 3.1. Classification accuracies of up to 85% for ancestry estimation and 98% for sex estimation were achieved using a multivariate approach. To measure the external validity and report more realistic performance statistics, an independent sample was tested. The postcrania from 180 black, white, and coloured South Africans were measured and classified using the custom postcranial database. A decrease in accuracy was observed for both ancestry estimation (79%) and sex estimation (95%) of the validation sample. When incorporating both sex and ancestry simultaneously the method achieved 70% accuracy, and 79% accuracy when sex-specific ancestry analyses were run. Classification matrices revealed that postcrania were more likely to misclassify as a result of ancestry rather than sex. While both sex and ancestry influence the size of an individual, sex differences are more marked in the postcranial skeleton and are therefore easier to identify. The external validity of the postcranial database was verified and therefore shown to be a useful tool for forensic casework in South Africa. While the classification rates were slightly lower than the original method, this is expected when a method is generalized.

**Keywords:** postcraniometric, multivariate, validation, Fordisc, classification

## **Introduction**

Acceptance of the *Daubert* criteria [1] as guidelines for best scientific practice acted as a paradigm shift in forensic research. Closer scrutiny was placed on both traditional methods and novel methods. Numerous papers were published regarding the role of the *Daubert* criteria in the field of anthropology and authors addressed the potential benefits that forensic anthropology would experience from conducting more specifically aimed validation studies [2-4]. While South African courts do not specifically require adherence to the standards recommended by the *Daubert* guidelines, similar criteria should be maintained in order to be on par with the international standards of best practice [5,6]. More recently, publication of reports by the National Academy of Sciences (NAS) and President's Council of Advisors on Science and Technology (PCAST) in the United States once again brought the issue of best practice to light, with both NAS and PCAST recommending further work

on methods demonstrating the continuous process of method testing and validation. [7,8]. As such, the validation of methodology is recognised as a key process in forensic anthropology and includes assessing population-specificity, testing the performance of methods on independent samples, and adding robust statistical tests to increase the power and reliability of analyses. Overall, method validation is required to adjust and refine methods for optimal performance within anthropological analyses, and forensic casework in particular. While most methods currently used in forensic casework have undergone some form of validation during method development – most commonly leave-one-out cross-validation – it is still essential to explore the external validity of the method and ensure older methods are still relevant to modern casework. As an example, the 1999 paper by İşcan and Steyn reported cross-validated classification rates of 98% for estimating ancestry from the crania and mandibles of modern black and white South Africans. Years later, L'Abbé et al. [9] published much lower correct classification rates (73%) for ancestry estimation in a contemporary sample of the same population groups. Numerous cases also presented as atypical with low posterior probabilities. The posterior probability refers to the likelihood of the individual belonging to a particular group, and the typicality probabilities refer to how typical the unknown individual is compared to members of the comparative groups [9]. When ancestry was estimated from skeletons with known demographic information, the 73% correct classification appears to be the more realistic result, as a substantial portion of the skulls misclassified when the İşcan and Steyn [10] formulae were applied. This example emphasizes the need for anthropological methods to be tested (and re-tested) to ensure accurate as well as realistic results when conducting forensic skeletal analyses of unknown remains. The cross-validated classification accuracy is important as this gives a reasonable estimate of performance on an unknown individual. However, there are other statistical parameters that inform the anthropologist if the classification is valid, such as posterior and typicality probabilities [11]. These probabilities give an indication of the strength and reliability of the derived discriminant function model. The validity and reliability of the models may play an important role in the final demographic estimates, particularly with parameters such as ancestry that demonstrate considerable group overlap and are prone to a degree of misclassification.

Recently, a postcranial database was created for use on modern South African remains. The original studies noted significant differences between the sexes as well as among the different population groups. Liebenberg et al. [12] observed an 85% cross-validated accuracy when assessing ancestry among black, white, and coloured South Africans, which is comparable to results obtained with the cranium [13]. Krüger et al. [14] achieved 98% correct classifications for sex estimation; this is similar to accuracies obtained with morphoscopic methods

applied to the pelvis [15,16]. The postcrania also outperformed both metric and morphoscopic methods used to identify sex from the skull [17-19]. Overall, the studies noted considerable size differences among the groups, with white South Africans displaying the largest measurement means and coloured South Africans the smallest. Black South Africans were found to be intermediate with substantial overlap with the coloured group, which typically resulted in decreased accuracies in comparison to white South Africans.

The availability of an alternative database to estimate biological parameters is particularly useful when popularly used elements, such as the skull or pelvis, are damaged, unavailable or provide inconsistent or ambiguous results. However, in order for the database to be confidently employed in practice, it must first undergo rigorous testing. The aim of the current study is to evaluate the validity of the postcranial South African database for sex and ancestry estimation using an independent sample.

## Materials and Methods

A total of 180 modern peer-reported black, white and coloured South Africans of known age, sex and ancestry were included in the sample (Table 1). Refer to Stull et al. [20] for more information on the origins and history of the South African groups. The skeletal material used in the study was obtained from the Pretoria Bone Collection (University of Pretoria) and the Kirsten Collection (Stellenbosch University). Even though both collections are comprised of individuals from the late 19<sup>th</sup> and 20<sup>th</sup> centuries, the material for the current study was limited to individuals born in the 20<sup>th</sup> century with birth years that range from 1900 to 1990. The skeletons originate from cadaveric material obtained by the respective institutions through donations or retrieval of unclaimed bodies [21,22]. Eleven postcranial bones were used, namely the clavicle, scapula, humerus, radius, ulna, sacrum, innominate, femur, tibia, fibula and calcaneus. Thirty-nine standard postcranial measurements were collected, to the nearest millimetre, following the measurement definitions by Moore-Jansen et al. [23]. All measurements were collected using a standard osteometric board, sliding caliper, and spreading caliper. Table 2 lists the measurements as well as the abbreviations used throughout the text.

**Table 1. South African postcranial test sample.**

| Group            | Abbreviation | Total Number (N) |
|------------------|--------------|------------------|
| Black Males      | BM           | 30               |
| Black Females    | BF           | 30               |
| White Males      | WM           | 30               |
| White Females    | WF           | 30               |
| Coloured Males   | CM           | 30               |
| Coloured Females | CF           | 30               |
|                  |              | <b>180</b>       |

**Table 2. Thirty-nine standard measurements and their associated abbreviations.**

|                                     |               |  |               |
|-------------------------------------|---------------|--|---------------|
| Clavicle maximum length             | <b>claxln</b> | Innominate height                          | <b>innoht</b> |
| Clavicle sagittal midshaft diameter | <b>claapd</b> | Iliac breadth                              | <b>iliabr</b> |
| Clavicle vertical midshaft diameter | <b>clavrd</b> | Femur maximum length                       | <b>femxln</b> |
| Scapula height                      | <b>scapht</b> | Femur bicondylar length                    | <b>fembln</b> |
| Scapula breadth                     | <b>scapbr</b> | Femur epicondylar breadth                  | <b>femebr</b> |
| Humerus maximum length              | <b>humxln</b> | Femoral head diameter                      | <b>femhdd</b> |
| Humerus epicondylar breadth         | <b>humebr</b> | Femur A-P subtrochanteric diameter         | <b>femsap</b> |
| Humeral head diameter               | <b>humhdd</b> | Femur transverse subtrochanteric diameter  | <b>femstv</b> |
| Humerus maximum diameter            | <b>hummx</b>  | Femur A-P midshaft diameter                | <b>femmap</b> |
| Humerus minimum diameter            | <b>hummw</b>  | Femur transverse midshaft diameter         | <b>femmtv</b> |
| Radius maximum length               | <b>radxln</b> | Tibia condylo-malleolar length             | <b>tibxln</b> |
| Radius A-P midshaft diameter        | <b>radapd</b> | Tibia proximal epiphyseal breadth          | <b>tibpeb</b> |
| Radius transverse midshaft diameter | <b>radtvd</b> | Tibia distal epiphyseal breadth            | <b>tibdeb</b> |
| Ulna maximum length                 | <b>ulnxln</b> | Tibia maximum diameter at nutrient foramen | <b>tibnfx</b> |
| Ulna dorso-volar diameter           | <b>ulndvd</b> | Tibia minimum diameter at nutrient foramen | <b>tibnft</b> |
| Ulna transverse diameter            | <b>ulntvd</b> | Fibula maximum length                      | <b>fibxln</b> |
| Ulna physiological length           | <b>ulnphl</b> | Fibula maximum diameter                    | <b>fibmdm</b> |
| Sacrum anterior height              | <b>sacaht</b> | Calcaneus maximum length                   | <b>calcxl</b> |
| Sacrum anterior breadth             | <b>sacabr</b> | Calcaneus middle breadth                   | <b>calcbr</b> |
| Transverse diameter of S1           | <b>sacs1b</b> |  |               |

Inter- and intra-observer reliability was assessed with repeated measurements taken from five individuals that were randomly selected from the sample. Technical error of measurement (TEM) was calculated to demonstrate the overall margin of error for each measurement. Additionally, the absolute technical error of measurement (%TEM) was calculated to assess the magnitude of measurement error in relation to the size of the measurement [24]. Bland-Altman plots were used to visually present the observer variation. All measurements collected per individual were uploaded into Fordisc 3.1 (FD3) and batch processed using the custom postcranial database created for South Africans as the reference sample [12,14]. FD3 is statistical software that employs discriminant analysis to classify skeletal remains into the most likely sex and ancestry groups by comparing measurements to a known reference database [25]. The FD3 program is a popular tool among many forensic anthropologists, as custom databases can be uploaded to run population-specific analyses. The software also has the added advantage of functioning on a case-specific basis, which enables the analysis of fragmentary or incomplete remains. Furthermore, FD3 provides extensive numerical results key to the evaluation of model performance and validity, which makes the method *Daubert* compliant [11,26]. None of the individuals included in the current study (hereafter referred to as the test sample) were used in the training sample to create the database. Separate analyses were run to classify the remains according to (1) sex (male/female with all populations pooled), (2) ancestry (black/white/coloured with sexes pooled), (3) sex and ancestry simultaneously (BM/BF/WM/WF/CM/CF), and (4) sex-specific ancestry. The sex-specific test was to gauge the effect of prior knowledge of sex on correctly identifying ancestry; the recorded sex was used to select only one sex from each

group. For example, if the test individual was noted to be male in the records, the individual was only compared to the BM, WM and CM reference groups in the analysis.

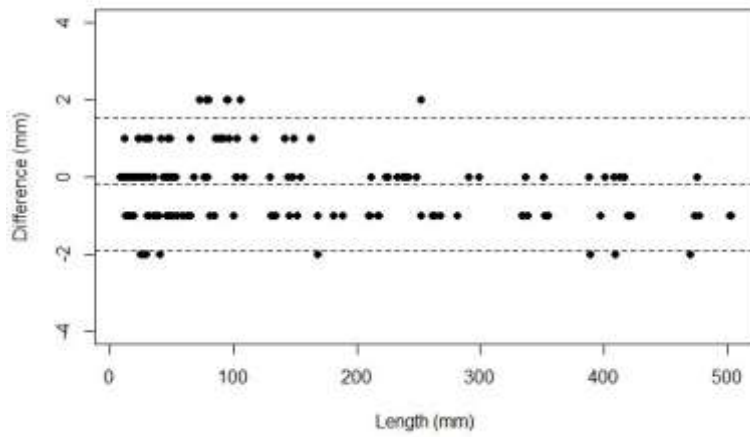
Stepwise variable selection was employed with each analysis. The stepwise procedure selects the most discriminatory variables during model creation and removes variables that do not contribute to the power of the model [27]. Following stepwise, the most accurate discriminant model is created using only the important variables, which helps maintain appropriate sample sizes and accurate calculations of parameters [28,29]. FD3 employs forward stepwise selection using Wilks' lambda criterion, so that all measurements are sequentially added to the model and the effects on the overall classification accuracy is gauged to determine if the variable will be included in the final classification model. The following criteria had to be adhered to in each FD3 analysis: (1) the sample size for each reference group had to be at least three times larger than the number of variables used; (2) the posterior probability had to be satisfactory, as a high posterior probability demonstrates greater confidence in a classification; and (3) typicality probabilities had to be larger than 0.05, or the individual would be labelled as atypical. The respective classifications were recorded for each individual and were used to calculate the percent correct, i.e. the total number of times the ancestry, sex, ancestry and sex simultaneously, and sex-specific ancestry were estimated correctly when compared to the known demographic information.

## **Results**

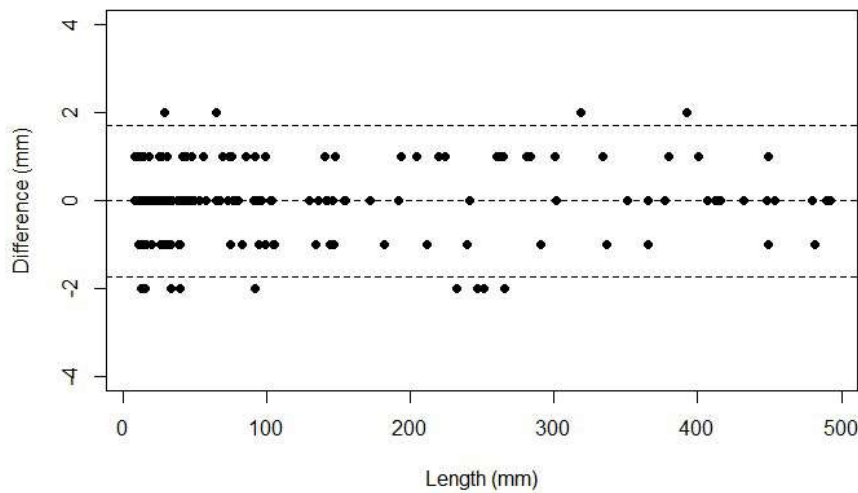
Observer agreement was assessed using TEM comparing measurements between two observers. The intra-observer TEM and %TEM ranged between 0 and 1.061, and 0% and 3.872%, respectively. Inter-observer agreement presented with slightly more variation, with the TEM and %TEM ranging between 0 and 1.183, and 0% and 7.246%, respectively (Table 3). The measurements that were found to be more variable include tibnft for the intra-observer analysis, and ulntvd, clavrd, and claapd for the inter-observer analysis. Bland-Altman plots demonstrated sufficient agreement, with none of measurements differing more than 2mm for either the intra- or inter-observer agreement analysis (Figures 1 and 2). As minimal measurement error was observed with the TEM and Bland-Altman plots, the observer agreement is considered satisfactory and all measurements were retained for further analyses.

**Table 3. Intra- and inter-observer error calculated using Technical error of measurement (TEM) and relative technical error of measurement (%TEM).**

|        | Intra-observer error |       | Inter-observer error |       |
|--------|----------------------|-------|----------------------|-------|
|        | TEM                  | % TEM | TEM                  | % TEM |
| claxln | 0,866                | 0,572 | 0,447                | 0,300 |
| clavrd | 0,316                | 3,131 | 0,548                | 5,889 |
| clapd  | 0,316                | 2,750 | 0,548                | 4,527 |
| scapht | 0,707                | 0,476 | 0,500                | 0,350 |
| scapbr | 0,707                | 0,695 | 0,548                | 0,559 |
| humxln | 0,548                | 0,170 | 0,775                | 0,241 |
| humebr | 0,632                | 1,076 | 0,707                | 1,192 |
| humhdd | 0,707                | 1,568 | 0,707                | 1,688 |
| hummx  | 0,447                | 2,090 | 0,316                | 1,589 |
| humw   | 0,447                | 2,662 | 0,447                | 2,727 |
| radxln | 0,775                | 0,319 | 0,837                | 0,332 |
| radapd | 0,447                | 3,666 | 0,447                | 3,727 |
| radtd  | 0,000                | 0,000 | 0,316                | 2,181 |
| ulnln  | 0,548                | 0,212 | 0,894                | 0,332 |
| ulndvd | 0,000                | 0,000 | 0,316                | 2,094 |
| ulntvd | 0,447                | 3,106 | 1,000                | 7,246 |
| ulnphl | 0,316                | 0,139 | 1,183                | 0,497 |
| sacaht | 0,894                | 0,872 | 0,500                | 0,508 |
| sacabr | 1,049                | 1,154 | 0,837                | 0,910 |
| sacs1b | 0,548                | 1,163 | 0,447                | 0,994 |
| innoht | 0,707                | 0,349 | 0,632                | 0,319 |
| iliabr | 0,447                | 0,315 | 0,316                | 0,217 |
| femxln | 0,548                | 0,119 | 0,447                | 0,098 |
| fembln | 0,837                | 0,184 | 0,316                | 0,069 |
| femebr | 0,775                | 0,988 | 0,316                | 0,419 |
| femhdd | 0,447                | 0,981 | 0,447                | 1,055 |
| femsap | 0,447                | 1,609 | 0,316                | 1,184 |
| femstv | 0,632                | 1,989 | 0,707                | 2,447 |
| femmap | 0,707                | 2,464 | 0,632                | 2,196 |
| femmtv | 0,000                | 0,000 | 0,447                | 1,720 |
| tibxln | 0,707                | 0,184 | 0,707                | 0,180 |
| tibpeb | 1,061                | 1,450 | 0,837                | 1,173 |
| tibdeb | 0,612                | 1,286 | 0,447                | 0,985 |
| tibnfx | 0,447                | 1,293 | 0,707                | 2,098 |
| tibnft | 0,949                | 3,872 | 0,775                | 3,050 |
| fibxln | 0,775                | 0,209 | 0,548                | 0,145 |
| fibmdm | 0,000                | 0,000 | 0,000                | 0,000 |
| calcbr | 0,707                | 1,837 | 0,000                | 0,000 |



**Fig 1** – Bland-Altman plot illustrating the intra-observer variation in measurement agreement.



**Fig 2** – Bland-Altman plot illustrating the inter-observer variation in measurement agreement.

Percent correct was used to gauge the predictive performance of the postcranial database. Table 4 presents the classification accuracies for the respective analyses, including sex, ancestry, sex-ancestry and sex-specific ancestry. The estimation of sex, when ancestry groups were pooled, revealed an overall accuracy of 95%. White males achieved the highest positive predictive values (100%), while coloured males achieved the lowest (90%). Separating the groups according to peer-reported ancestry (with the sexes pooled for each group) reveals an overall accuracy of 79%. White South Africans demonstrated the greatest separation from the other groups and achieved the highest accuracy (95%). Black and coloured South Africans demonstrated substantial group overlap, as shown by the misclassifications observed in the classification matrix (Table 5). Following the large amount of overlap, both groups have decreased accuracies compared to the white group, with black South Africans obtaining the lowest accuracy (70%).



**Table 4. Overall classification accuracies for the respective analyses using the postcranial South African database.**

|                       |     |
|-----------------------|-----|
| Sex                   | 95% |
| Ancestry              | 79% |
| Sex-Ancestry          | 70% |
| Sex-specific Ancestry | 79% |

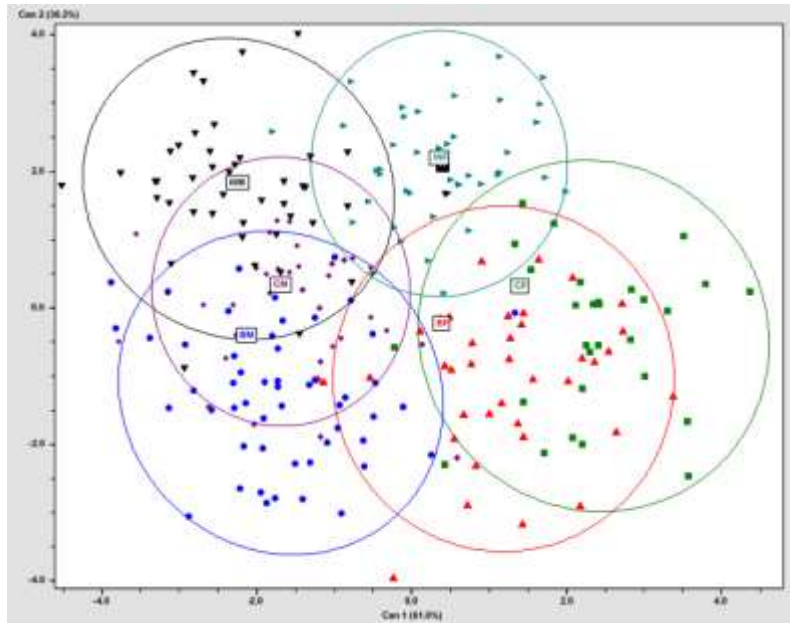
**Table 5. Classification matrix showing classification of groups according to ancestry.**

| <u>Group:</u> |          | <u>Classifies into:</u> |          |          | <i>n</i> | % Correct |
|---------------|----------|-------------------------|----------|----------|----------|-----------|
|               |          | <b>B</b>                | <b>W</b> | <b>C</b> |          |           |
|               | <b>B</b> | 42                      | 4        | 14       | 60       | 70%       |
|               | <b>W</b> | 0                       | 57       | 3        | 60       | 95%       |
|               | <b>C</b> | 12                      | 4        | 44       | 60       | 73%       |

When assessing sex and ancestry simultaneously (i.e. the sex-ancestry analysis), the sample demonstrated a percent correct of 70%. While an accuracy of 70% seems low, it should be noted that classification is into one of six groups; as such, 70% is more than 50% greater than chance (chance is calculated as 16.7%). White males and females achieved the highest accuracies with 87% each, whereas coloured males achieved the lowest accuracy with 47%. The classification matrix (Table 6) shows misclassification according to ancestry rather than sex. For example, a female still classifies as a female, but as a female from a different population group. This is true for all groups, except for the coloured male South Africans, where numerous sex misclassifications were noted in addition to ancestry misclassifications. Figure 3 demonstrates the within- and between-group variation present within the sample, which shows greater overlap among the sexes. As such, members of the same sex are more likely to misclassify into a different ancestry group rather than the opposite sex of the same ancestry group (e.g. a black female would more likely misclassify as a coloured or white female rather than misclassify as a black male). The ellipses also reveal substantial overlap between both sexes of the black and coloured groups, indicating that black and coloured individuals are more likely to misclassify than for either group to misclassify as white. White South Africans demonstrate less overlap with the other groups; however, white males and females are still more likely to misclassify according to ancestry (into an incorrect population group) rather than misclassifying according to sex (into the incorrect sex).

**Table 6. Classification matrix showing classification of groups according to sex and ancestry.**

|               |           | Classifies into: |    |    |    |    |    | <i>n</i> | % Correct |
|---------------|-----------|------------------|----|----|----|----|----|----------|-----------|
|               |           | BF               | BM | CF | CM | WF | WM |          |           |
| <b>Group:</b> | <b>BF</b> | 18               | 1  | 10 | 0  | 1  | 0  | 30       | 60%       |
|               | <b>BM</b> | 3                | 18 | 0  | 5  | 0  | 4  | 30       | 60%       |
|               | <b>CF</b> | 5                | 0  | 23 | 2  | 0  | 0  | 30       | 77%       |
|               | <b>CM</b> | 2                | 11 | 2  | 14 | 0  | 1  | 30       | 47%       |
|               | <b>WF</b> | 1                | 0  | 2  | 0  | 26 | 1  | 30       | 87%       |
|               | <b>WM</b> | 0                | 1  | 0  | 3  | 0  | 26 | 30       | 87%       |



**Fig 3** – Canonical variate plot demonstrating group variation and overlap among the sex-ancestry cohorts of modern South Africans when using the postcranial skeleton.

Taking the effects of sex into account through the use of sex-specific ancestry analyses achieve combined correct classifications up to 79%. Similar to the results obtained with the sex-ancestry analysis, black and coloured South Africans are more likely to misclassify with each other than misclassifying as white South Africans (Table 7). While the accuracies for ancestry estimation and for sex-specific ancestry (requiring a sex estimate prior to estimating ancestry) are similar, sex-specific ancestry estimation is more consistent among the groups. Where coloured males only classified into the correct ancestry groups in 53% of the analyses, the coloured females classified into the correct ancestry group in 93% of the analyses. The large disparity between the sexes is decreased when sex-specific ancestry is estimated, where the ancestry is correctly estimated for 73% of coloured males and 80% of coloured females. Similarly, in black South Africans, black females only classify correctly according to ancestry in 60% of the analyses, whereas black males classify correctly in 80% of the

analyses. With sex-specific ancestry, black females classify into the correct ancestry group in 70% and black males in 67% of the analyses.

**Table 7. Classification matrix showing classification of groups according to sex-specific ancestry.**

| <u>Group:</u> | <u>Classifies into:</u> |    |    |    |    |    | <i>n</i> | % Correct | % Correct |
|---------------|-------------------------|----|----|----|----|----|----------|-----------|-----------|
|               | BF                      | BM | CF | CM | WF | WM |          |           |           |
| BF            | 21                      | -  | 8  | -  | 1  | -  | 30       | 70%       | 68.5%     |
| BM            | -                       | 20 | -  | 7  | -  | 3  | 30       | 67%       |           |
| CF            | 5                       | -  | 24 | -  | 1  | -  | 30       | 80%       | 76.5%     |
| CM            | -                       | 7  | -  | 22 | -  | 1  | 30       | 73%       |           |
| WF            | 1                       | -  | 1  | -  | 28 | -  | 30       | 93%       | 93%       |
| WM            | -                       | 1  | -  | 1  | -  | 28 | 30       | 93%       |           |

## Discussion

This validation study demonstrated a decrease in accuracy for postcranial sex and ancestry estimation compared to the original publications; however, the accuracies are still comparable to that of published methods frequently employed and recognized as best practices in forensic anthropology [12,14]. Even though the current study achieved lower classification rates, the findings were expected due to the fact that it was an independent sample. In agreement with published literature [30], postcranial elements are useful for generating sex estimates and outperforms both metric and morphoscopic methods employing the cranium (85-86% and 93%, respectively) [17-19]. The postcraniometric database obtained classification rates similar to ancestry estimation using the cranium (73-84%) [9,13]. The classification matrices revealed that the postcranial database is more likely to misclassify an individual according to ancestry rather than sex; this is in contrast to data on the cranium, which demonstrated that individuals tend to misclassify according to sex [9]. A possible explanation for this result is that craniometric variation is a product of both size and shape, while postcraniometric variation is mainly influenced by size. Both sex and ancestry influence the overall size of an individual; however, the sexual size dimorphism appears to be more marked in the postcranial skeleton and as such is easier to identify (thus obtaining greater classification accuracies). While the interaction with sexual size dimorphism masks more subtle variation attributable to ancestry, resulting in misclassification into incorrect ancestry groups. In other words, overlap among the groups is observed as a result of size differences attributable to both sex and ancestry. For instance, black males and coloured males demonstrate substantial overlap and tend to misclassify as one another. Although black and coloured South Africans overlap, significant differences have been observed between the groups [12]; thus, a greater likelihood for black males and black females to misclassify is expected. However, because of marked sex differences and black and coloured groups being similar in size, black males

more frequently misclassify as coloured males, and black females misclassify as coloured females. The results suggest that postcrania are better at identifying sex than ancestry.

Osteometric variation has been shown to demonstrate correlations with population history and structure [20,31]. The results of the current study are consistent with previously published postcraniometric data [12,14]. Specifically, white South Africans were found to be fairly homogenous with considerable separation from the other population groups, while coloured South Africans demonstrate the widest range of within-group variation. The coloured males in particular are highly variable, resulting in poor classification accuracies (47%). The different patterns of variation among coloured South Africans, as documented in genetic literature and demonstrated with craniometric data, has been ascribed to sex-biased gene flow. In accordance with historical records, mitochondrial DNA (mtDNA) exhibit a substantial Khoe-San contribution, while the Y-chromosome shows pronounced contributions from Europeans and Bantu-speakers [20,32,33]. The wide-ranging contributions to the genetic make-up of coloured males result in greater levels of variation, observed as great size variation in the postcranial skeleton. Greater variation results in more group overlap, ultimately leading to increased misclassifications. However, when conducting sex-specific ancestry analyses the effect of population-specific sexual dimorphism is removed and the model only needs to assess size differences attributable to ancestry. This is useful for groups, such as coloured South African males, who not only misclassified according to ancestry, but also misclassified according to sex. Therefore, prior knowledge of sex can increase the accuracy by decreasing group overlap and more effectively separating the groups.

Another possible explanation for the drop in accuracy observed with the current study is the sample itself. First, validation studies are expected to produce lower accuracies, as independent samples (with potential different levels of variation) are used. Second, the reference samples used during method creation make use of “clean data”. One of the steps of the data cleaning process involves the identification and removal of outliers. Outliers may be present for several reasons such as recording errors, mis-measurement or underlying pathological conditions. However, outlying values may also be the result of individuals with a valid, albeit unique combination of characteristics identifying them as different from the other individuals within the same group [34]. The removal of outliers is typically recommended as multivariate techniques, like discriminant analysis, may be drastically affected by their presence. Unfortunately, it is difficult to determine if skeletal measurements collected from forensic material contain any potentially outlying or extraordinary values, especially when assessed in multivariate combination. To this effect, an individual may misclassify as a result of

a unique combination of variables. When subsequently running measurements from a forensic case against reference databases in FD3 poor model accuracies, low posterior probabilities and occasionally non-significant typicality probabilities may be observed. Some of the issues encountered with assessing data from unknown remains (such as potential outliers and different variance structures) may be lessened by only using variables with one standard deviation away from the group means. By removing any variables with greater standard deviations the amount of variation associated with the unknown individual is decreased; this will allow the measurement means of the individual to appear more typical in comparison with the members of the reference groups and thereby making a correct classification more likely.

The postcranial database has been shown to be valid and reliable and can be used to estimate sex and ancestry in forensic anthropological casework in South Africa. The database can be used through the application of FD3 or R [35].

## References

1. *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579 (1993).
2. Christensen AM (2004) The impact of Daubert: Implications for testimony and research in forensic anthropology (and the use of frontal sinuses in personal identification). *J Forensic Sci* 49:427-430.
3. Grivas CR, Komar DA (2008) *Kumho, Daubert* and the nature of scientific inquiry: Implications for Forensic Anthropology. *J Forensic Sci* 53:771-776.
4. Christensen AM, Crowder CM (2009) Evidentiary standards for Forensic Anthropology. *J Forensic Sci* 54:1211-1216.
5. Allan A, Louw DA (2001) Lawyers' perception of psychologists who do forensic work. *S Afr J Psychol* 31:12-20.
6. Meintjes-Van der Walt L (2003) The proof of the pudding: The presentation and proof of expert evidence in South Africa. *J Afr Law* 47:88-106.
7. Committee on Identifying the Needs of the Forensic Sciences Community, National Research council (2009) Strengthening forensic science in the United States: A path forward.
8. President's Council of Advisors on Science and Technology (2016) Report to the president Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods, Washington DC.

9. L'Abbé EN, Kenyhercz MW, Stull KE, Keough N, Nawrocki S (2013) Application of Fordisc 3.0 to explore differences among crania of North American and South African blacks and whites. *J Forensic Sci* 6:1579-1583.
10. İşcan MY, Steyn M (1999) Craniometric determination of population affinity in South Africans. *Int J Legal Med* 112:91-97.
11. Ousley SD, Jantz RL (2012) FORDISC 3 and Statistical Methods for Estimating Sex and Ancestry. In: Dirkmaat DC (ed) *A companion to Forensic Anthropology*. Blackwell Publishing LTD, pp 311-329.
12. 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.
13. Stull KE, Kenyhercz MW, L'Abbé EN (2014) Ancestry estimation in South Africa using craniometrics and geometric morphometrics. *Forensic Sci Int* 245:206e1.
14. Krüger GC, L'Abbé EN, Stull KE (2017) Sex estimation from the long bones of modern South Africans. *Int J Legal Med* 131:275-285.
15. Kales AR, Ousley SD, Vollner JM (2012) A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *Am J Physical Anthropol* 149: 104-114.
16. Kenyhercz, MW, Kales AR, Stull KE, McCormick KA, Cole S J (2017) Worldwide Population Variation in Pelvic Sexual Dimorphism: A Validation and Recalibration of the Kales et al. Method. *Forensic Sci Int* 277:259.e1-259.e8.
17. Steyn M, İşcan, MY (1998) Sexual dimorphism in the crania and mandibles of South African whites. *Forensic Sci Int* 98: 9-16.
18. Dayal MR, Spocter MA, Bidmos MA (2008) An assessment of sex using the skull of black South Africans by discriminant function analysis. *HOMO* 59: 209-221.
19. 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.
20. Stull KE, Kenyhercz MW, L'Abbé EN, Tuamsuk P (2016) The Craniometric Implications of a Complex Population History in South Africa. In: Pilloud MA, Hefner JT (eds) *Biological Distance Analysis: Forensic and Bioarchaeological Perspectives*. Elsevier Inc, pp 245-263.
21. L'Abbé EN, Loots M, Meiring JH (2005) The Pretoria Bone Collection: A modern South African skeletal sample. *Homo* 56:197-205.

22. Alblas A, Greyling LM, Geldenhuys E (2018) Composition of the Kirsten Skeletal Collection at Stellenbosch University. *S Afr J Sci* 114:e1-e6.
23. Moore-Jansen PM, Ousley SD, Jantz RL (1994) Data collection procedures for forensic skeleton material. The University of Tennessee, Knoxville: Department of Anthropology. p. 70-82.
24. Stomfai S, Ahrens W, Bammann K, Kovács É, Mårild S, Michels N, Moreno LA, Pohlabein H, Siani A, Tornaritis M, Veidebaum T, Molnár D (2011) Intra- and inter-observer reliability in anthropometric measurements in children. *Int J Obesity* 35:45-51.
25. Jantz RL, Ousley SD (2005) *FORDISC 3.0: Personal Computer Forensic Discriminant Functions*. Knoxville: University of Tennessee.
26. Ousley SD, Jantz RL (2013) *FORDISC 3: Third generation of computer-aided forensic anthropology*. *Rechtsmedezin* 23:97-99.
27. Kachigan SK (1991) *Multivariate statistical analysis: A conceptual introduction*. Radius Press, pp 216-235.
28. Huberty C (1994) *Applied discriminant analysis*. New York: Wiley and sons.
29. Tabachnick BG, Fidell LS (2007) *Using multivariate statistics* 5<sup>th</sup> ed. Boston: Pearson Education.
30. Spradley MK, Jantz RL (2011) Sex Estimation in Forensic Anthropology: Skull Versus Postcranial Elements. *J Forensic Sci* 56: 289-296.
31. Relethford JH (2009) Race and global patterns of phenotypic variation, *Am J Phys Anthropol* 139:16-22.
32. Patterson N, Petersen DC, van der Ross RE, Sudoyo H, Glashoff RH, Marzuki S, Reich D, Hayes VM (2010) Genetic structure of a unique admixed population: Implications for medical research. *Hum Mol Gen* 19:411-419.
33. Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C 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.
34. Hair JF, Black WC, Babin BY, Anderson R, Tatham R (2010) *Multivariate Data Analysis: A Global Perspective*. Pearson Prentice Hall.
35. R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.