Osteometric sorting of skeletal elements from a sample of modern Colombians
Rodríguez, Juan Manuel Guerrero; Hackman, Lucina; Martínez, Wendy; Medina, César Sanabria

Published in: International Journal of Legal Medicine

DOI: 10.1007/s00414-015-1142-1

Publication date: 2016

Document Version
Peer reviewed version

Link to publication in Discovery Research Portal

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Osteometric sorting of skeletal elements from a sample of modern Colombians

Rodríguez, Juan Manuel Guerrero; Hackman, Sarah; Martínez, Wendy; Medina, César Sanabria

Published in:
International Journal of Legal Medicine

DOI:
10.1007/s00414-015-1142-1

Publication date:
2016

Document Version
Peer reviewed version

Link to publication in Discovery Research Portal

Citation for published version (APA):

The final publication is available at Springer via http://dx.doi.org/10.1007/s00414-015-1142-1
OSTEOMETRIC SORTING OF SKELETAL ELEMENTS FROM A SAMPLE OF MODERN COLOMBIANS: A PILOT STUDY

Juan Manuel Guerrero Rodríguez1 (MSc); Lucina Hackman1* (PhD); Wendy Martínez2, César Sanabria Medina3 (PhD).

ABSTRACT: The Colombian armed conflict has been catalogued not only as the longest civil war in the western hemisphere, but also as having one of the highest indexes of missing persons. Among the several challenges faced by forensic practitioners in Colombia, the commingling of human remains has been recognised as one of the most difficult to approach. The method of osteometric sorting described by Byrd & Adams and Byrd (2008) has proven relevant as a powerful tool to aid in the re-association process of skeletal structures. The aim of this research was to evaluate the three osteometric sorting models developed by Byrd (2008) (paired elements, articulating bone portions and other bone portions) in a sample of modern Colombian individuals. A set of 39 linear measurements were recorded from a sample of 100 individuals (47 females and 53 males aged between 20 and 74 years, and 18 and 77 respectively), which were used to create a reference sample database. A different subset of 8 individuals (5 females aged between 23 and 48 years, and 5 males aged between 27 and 43 years) was employed to randomly create eight small-scale commingled assemblages for the purposes of testing the osteometric sorting models. Results demonstrate that this method has significant potential for use in the Colombian forensic context.

Keywords: Forensic Anthropology, Commingled Human Remains, Osteometric Sorting, Armed Conflict, Colombia.

1 Centre for Anatomy and Human Identification, University of Dundee (MSI/WTB/JBC Complex, Dow Street - DD1 5EH Dundee, Scotland).
2 Departamento de Antropología, Universidad Nacional de Colombia (Avenida Carrera 30 # 45- Bogotá, Colombia).
3 Instituto Nacional de Medicina Legal y Ciencias Forenses / Facultad de Medicina, Universidad Antonio Nariño – (Carrera 3 este # 47A-15 - Bogotá, Colombia).
* Corresponding author additional details: (Tel: 01382 386311 – email: l.hackman@dundee.ac.uk)
Introduction

When analysing human skeletal remains, forensic anthropologists worldwide make use of several techniques to obtain the biological profile of an individual in question (sex, stature, age-at-death, ancestry, pathology, etc.) to aid in the process of personal identification [1, 2]. The effectiveness and evidentiary strength of these methods is improved when they are used on the complete skeleton of the individual. Unfortunately, this ideal scenario is not always possible in forensic contexts, since many factors can disrupt the integrity of a human skeleton. One of these is the commingling of human remains, which is defined as the presence of duplicate skeletal elements from 2 or more individuals in the same assemblage [3] with no apparent indication as to which individual they belong. This issue can severely interfere with the personal identification process limiting the anthropologist in their ability to obtain and synthesize data from multiple parts of the skeleton; eventually it can delay the return of the remains to the next of kin and the investigative process in general [4]. It has been noted that this type of situation tends to be a common component of specific types of incidents, such as mass fatality events and mass graves resulting from armed conflict contexts [5]. In Colombia, commingling of skeletal elements has been reported to commonly occur as a result of the dismemberment of persons and subsequent disposal in collective burials (practice usually attributed to paramilitary groups), poor registry and deposition of unidentified individuals in unmarked collective graves within cemeteries, and bombings resulting from military operations and acts of terrorism [6, 7].

Scientific literature has focused little attention on the issues of commingling of human remains (8, 9), when compared with other aspects of forensic anthropology. As a general overview, the published material on this subject usually presents descriptions of analytical procedures and best practices using case studies as examples, or provides results of experimental research that test different methods than can aid to re-associate mixed assemblages. Other relevant academic literature that contributes to the understanding of this subject is provided from experiences gathered from archaeological contexts. A detailed and comprehensive review of some of this literature has been previously presented by Adams & Byrd [10] and Ubelaker [11], and it will not be further discussed in this paper.

Among the several sorting techniques available to resolve the commingling of human remains, the osteometric sorting method described and developed in detail by Byrd & Adams [12] and Byrd [4] stands as a promising tool to address this issue. This method makes use of statistical models built upon measurement data from a large reference sample composed of pooled individuals representing both sexes and several racial groups, to formally compare the relationships between size and shape of the skeletal elements. In their study, Byrd and Adams [12] have proposed that the segregation decisions are made by testing the statistical null hypothesis that two specimens, given their size and shape, could have originated from a single individual. As a result, the segregation of the remains is based on a solid statistical base that removes subjective judgment decisions. More recently, Byrd [4] presented three approaches or models of osteometric sorting: 1) comparison of left and right bones by emphasizing their shape; 2) comparison of adjoining bones based on the
correlation of corresponding regions at joints; and 3) the comparison of the sizes of cones with the use of regression models.

Similar to other anthropological techniques that rely on statistical models for hypothesis testing, the reference data on which these models are grounded must represent the population of the individuals that are subject to analysis. The previous studies that have evaluated the osteometric sorting method, have based their reference data sample on measurements from American Whites, American Blacks, Asians and a few Mexican individuals [4]. As a result, if this method is used on different populations without previous testing, it can lead to possible bias and incorrect segregation of the remains.

Given the great potential that the osteometric sorting method provides for the separation of commingled assemblages of human remains, the main objective of this research was to evaluate the models and findings developed in previous studies [4, 12] in a sample of modern Colombian individuals in order to assess if this method can be added to the toolkit of techniques used by the forensic anthropologists of this country for the analysis of their daily casework.

Materials and Methods

Sample

For the purposes of this research, a total sample of modern Colombians composed of 52 female individuals aged between 20 and 74 years (x = 41 years; standard deviation (S.D) = 14 years) and 58 male individuals aged between 18 and 77 years (x = 36 years; standard deviation (S.D) = 15 years) was used. These individuals belong to the Skeletal Reference Collection from Colombian Population (Colección Ósea de Referencia de Población Colombiana), currently held and curated by the National Institute of Legal Medicine and Forensic Sciences (NILMFS) in Bogotá, since the year 2010. The individuals of this modern skeletal collection have documented information of date of death, sex, age at death, stature, place of birth, and in some cases, their cause/manner of death. In addition, all the remains of the skeletal collection were recovered from burial contexts by the NILMFS once they had reached the maximum burial period allowed in the public cemeteries (4 years). In accordance with Colombian laws and customs, after this period expires, the bodies must be reclaimed by a next-of-kin member; otherwise they will be cremated and disposed in a common space in the cemetery.

In order to fulfil the requirements of the osteometric sorting method, a reference sample of 100 individuals (47 females and 53 males aged between 20 and 74 years, and 18 and 77 respectively) derived of the total sample was employed. The remaining 8 individuals (5 females aged between 23 and 48 years, and 3 males aged between 27 and 43 years), were used to randomly create small-scale commingled cases to test the osteometric sorting models which were created from analysis of measurements taken from the original 100
individuals. These cases were created by using the measurements of the skeletal elements and did not involve the mixing of any skeletal elements.

**Variables and data collection**

To create an adequate database to support the osteometric sorting models, a set of 50 linear measurements from postcranial elements were recorded in millimetres for each one of the individuals of the sample (Table 1 - Appendix 1). All of the measurements were taken from modern research literature and are recognized anthropometric measurements. The selection of the measurements for this research considered various criteria, including: their proven usefulness and minimal intra and inter-observer error documented by previous studies [12], their potential to represent the size and shape of each element and their potential for their use in isolation.

In order to develop a reliable reference data sample that can be used in the future by the Colombian forensic anthropologists and other investigators, quality control measures were taken during the collection of the data. The measurements from each skeleton were taken only when the preservation of its structures was good in order to ensure that taphonomic influences would not alter the quality of the measurements. Additionally, variables that were potentially altered either due to pathological and/or traumatic conditions were disregarded.

The gathering of data was conducted by the first author and a trained assistant approved by the NILMFS, whose work was guided and supervised at all times. Intra-observer error in the data collected by the first author was carried out by measuring a set of 4 individuals (2 females and 2 males) one time per week during a period of three weeks. Inter-observer error of the data collected by this author and his assistant was tested by comparing the measurements obtained from the same set of 4 individuals by both researchers.

**Statistical procedures**

The database created by these measurements was carefully scanned for outliers. When detected, a detailed evaluation to assess if they could have been attributed to data entry, measurement error or extreme values, was conducted. Only when outliers were caused by any of the first two causes, these were corrected when possible, or deleted if the detected error was uncorrectable. Extreme values where not removed from the dataset, since under normal circumstances the appearance of these types of data is expected. Shapiro-Wilk test was conducted for each one of the evaluated variables to assess if the data have a considerable deviation from the normal distribution and to understand the effects of the extreme values in the dataset.

Furthermore, to test the intra and inter-observer reliability of the measurements, intra-class correlation coefficients (ICC) were calculated based on the suggestions discussed by Ferrante & Cameriere [17]. ICC were computed using a two factor mixed effects model and type absolute agreement.
The measured variables of the subsample of 10 individuals (5 females and 5 males) were used to artificially create 6 commingled assemblages to test the overall effectiveness of the osteometric sorting models designed by Byrd [4], and verify their performance on this sample. The selected elements on each one of the assemblages were selected by considering possible forensic scenarios where each osteometric sorting model could be used. The same individuals were not compared twice in the same test application. Statistical calculations for these tests were performed in the software Microsoft Excel. Details of each test case and statistical procedures for each model application will be presented on the following sections:

**Osteometric sorting model #1 – Paired elements:**

This model is aimed to compare paired bones (left and right) by emphasizing their shape. To test this model, a set of partial remains from 4 individuals was used to create 2 different scenarios (Table 1). In all these tests, it is assumed that the represented skeletal structures are complete.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Sex</th>
<th>Age</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>M</td>
<td>33</td>
<td>Right and left humeri, right ulna, right and left radii.</td>
</tr>
<tr>
<td>B</td>
<td>M</td>
<td>43</td>
<td>Left humerus, left radius, right and left ulnae.</td>
</tr>
<tr>
<td>Test 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>F</td>
<td>41</td>
<td>Right and left femora, left tibiae, right and left fibulae</td>
</tr>
<tr>
<td>D</td>
<td>F</td>
<td>48</td>
<td>Right femur, right and left tibiae, left fibulae</td>
</tr>
</tbody>
</table>

**Table 1. Osteometric sorting model #1. Tests and sample characteristics.**

For all tests of the osteometric model #1, all the selected measurements from every bone were used. To test the model, the formula proposed by Byrd [4] was used:

\[ D = \Sigma(a_i - b_i) \]

In this equation, *a* is the right side bone measurement(s) *i*, and *b* is the left side bone measurement(s) *i* for each of the variables that were compared. Once the values of *D* were obtained, these were compared against “0”, since this represent the null hypothesis of no difference and using the standard deviation of *D* in the reference sample (n=100). The deviations of “0”, divided by the reference sample standard deviation was compared against the two-tailed t-distribution with *N*-1 degrees of freedom (*df*) to obtain *p-values*. A level of significance of 0.10 was used in these tests.
The test 1 is formed by an artificially commingled set of skeletal elements of the upper limb obtained from 2 male individuals (A & B). The measurement values for each one of the variables is presented in the Table 1 of the Appendix 2. Moreover, the Test 2 included skeletal remains of the lower limb from two female individuals labeled as C and D respectively. The measurement values of the skeletal elements are presented on the Table 2 of the Appendix 2.

**Osteometric sorting model # 2 – Articulating bone portions:**

To test this model, skeletal elements from 4 individuals were used to create 2 different commingled scenarios (Table 2). Since this model is designed to evaluate adjoining bone portions, only measurements from these areas were employed in this study.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Sex</th>
<th>Age</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>23</td>
<td>Left innominate, left femur, right and left tibiae, and left talus.</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>27</td>
<td>Right and left femora, right and left tibiae and left talus.</td>
</tr>
<tr>
<td>Test 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>M</td>
<td>33</td>
<td>Right scapula, right humeri, and left ulna.</td>
</tr>
<tr>
<td>H</td>
<td>F</td>
<td>32</td>
<td>Right and left humeri, and left ulna.</td>
</tr>
</tbody>
</table>

**Table 2. Osteometric sorting model # 2. Tests and sample characteristics.**

Osteometric model #2 was tested by using the differences in sizes in the articular portions of adjoining bones. For instance, a right femur and a right tibia can be compared by subtracting the bicondylar breadth of the distal femur from the proximal epiphyseal breadth of the tibia. To test the model, the equation presented by Byrd [4] was used:

\[ D = c_i - d_j \]

In this formula, the measurement \( i \) from bone \( c \) is subtracted from the measurement \( j \) of bone \( d \). Once obtained the \( D \) value for the individuals in the tests, the mean \( D \) value from the reference sample was calculated. Subsequently, the null hypothesis that two specimens have an appropriate size to have been originated in one individual [4] was tested by comparing the previously obtained \( D \) values. Then, the deviations of \( D \) from the reference sample means, divided by the reference sample standard deviations were evaluated against a two-tailed \( t \) distribution to obtain a \( p-value \). The 0,05 significance level was used in these tests.
In Test 3, measurements from adjoining elements of the hip joint, knee joint and ankle joint, were collected from two female individuals labeled as E and F. Additionally, to develop the Test 4, skeletal measurements from the shoulder and elbow joint were collected from one female (H) and one male individual (G). The skeletal measurements employed for the respective osteometric comparisons of the Test 3 and 4 are shown at the Tables 3 and 4 of the Appendix 2.

Osteometric sorting model # 3 – Other bone portions:

To compare sizes and shapes of different bones, this study used the measurements of skeletal elements from 4 individuals to create 2 different commingled assemblages (Table 3). As demonstrated by Byrd & Adams [XIII] and Byrd [IV], the best index to make this comparison can be obtained through a linear combination. In that sense, this study used measurements from the girth, breadth and length of several bones to create a natural logarithm by summing these values. The result of this sum was used to create linear regression models. The reason to not include maximum lengths in some tests was motivated by the fact that the recovery of incomplete and fragmented remains is a common issue on both forensic and archaeological scenarios. Nevertheless, it is important to point out that according to the results presented by Byrd [IV], the linear models based on these measurements can perform nearly as well as the ones that only include lengths.

<table>
<thead>
<tr>
<th>Individual</th>
<th>Sex</th>
<th>Age</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>F</td>
<td>27</td>
<td>Right humerus and right radius</td>
</tr>
<tr>
<td>J</td>
<td>F</td>
<td>48</td>
<td>Right humerus and right radius</td>
</tr>
<tr>
<td>K</td>
<td>M</td>
<td>27</td>
<td>Left humerus, left femur</td>
</tr>
<tr>
<td>L</td>
<td>F</td>
<td>27</td>
<td>Left humerus, left femur</td>
</tr>
</tbody>
</table>

Table 3. Osteometric sorting model #3. Tests and sample characteristics.

In addition, according to the suggestions of Byrd [IV], the \( t \)-value used for comparison against the \( t \) distribution with N-2 degrees of freedom to test the null hypothesis of association was calculated by using the following model:

\[
t = \left| y^* - y_i \right| \times \left[ \left( \frac{1}{N} \right) \times \left( \frac{x_i - x}{N \cdot S^2_e} \right) \right]
\]

In this equation, \( y^* \) is the predicted value from the regression model, \( y_i \) is the dependant variable value of the case specimen, \( S^e \) is the regression model standard error, \( x_i \) is the independent variable value of the case.
specimen, \( x \) is the reference sample mean for the independent variable, and \( S_x \) is the reference sample standard deviation of the independent variable. A level of significance of 0.10 was used in these tests.

For the Test 5, a comparison was made between two right humeri and two right radii obtained from two female individuals who were designated as “I” and “J” respectively. A natural logarithm was created from the sum of two girth measurements of each structure, which was subsequently used to create a linear regression model (Table 5 – Appendix 2). Conversely, for the Test 6, comparison of two left humeri (HUM(K) and HUM(L)) and two left femora (FEM(K) and FEM(L)) obtained from one female (L) individual and one male (K) individual, was carried out. Only length measurements were obtained for comparison (Table 6 – Appendix 2).

**Results**

**Assessment of normality**

For all the variables evaluated with the Shapiro-Wilk test, the \( p \)-values obtained where found to be non-significant (\( p > 0.05 \)). This demonstrates that the distribution of the data is not significantly different from a normal distribution. This also proves that the extreme values have not created a significant impact on the distribution of the dataset.

**Assessment of intra and inter-observer reliability**

The intra-observer and inter-observer reliability between the set of recorded measurements obtained from the 4 individuals was found to be high (ICC = 1; 95% confidence Interval from 1 to 1).

**Test applications of osteometric sorting models**

**Osteometric sorting model # 1 – Paired elements**

After following the procedures listed previously, statistical values were obtained for each one of the comparisons. Details of each one of the statistical calculations are provided in the Tables 1 to 3 (Test 1) and Tables 4 to 6 (Test 2) of the Online Resource 1.

**Test 1**

For each one of the comparisons made within this test, two possible courses of action were possible. The first one dictates that the skeletal elements could be segregated into their respective individuals, whereas the second one indicates that they could be segregated into their opposite individuals. The statistical calculations revealed that when skeletal elements of the
same individuals where compared, the \( p \)-value was always greater than 0.10, indicating a clear acceptance of the null hypothesis. On the contrary, when opposed individuals where compared the \( p \)-value indicated a rejection of the null hypothesis (\( p < 0.10 \)). According to this, the most reasonable course of action to follow is the first one, proving that in this test the segregation of the skeletal elements was correct.

Test 2

Statistical calculations of this test revealed that for all of the comparisons made, the null hypothesis is accepted (\( p > 0.10 \)). However, it must be noted that according to the \( p \)-values obtained (Tables 4 to 6 – Online Resource 1), the probability of association between the skeletal elements and their corresponding individual is slightly higher than the probability calculated for when they are compared against their opposite individual. Nevertheless, in this case, a segregation based solely on the results of the osteometric model is not recommended, since the results confirm that there is a poor differentiation between the two individuals. This last consideration can be explained based on the similarities in size and shape of the skeletal elements measured, and the age of the individuals in question.

Osteometric sorting model # 2 – Articulating bone portions

The statistical values obtained for all the comparisons made to test this osteometric sorting model are listed on the Tables 7 to 9 (Test 3) and Tables 10 and 11 of the Online Resource 1.

Test 3

The results for the osteometric comparisons reveal different courses of action for each one of them. Firstly, for the evaluation of elements at the hip joint, the null hypothesis was accepted for both comparisons (\( p > 0.05 \)). Although the probability of association between ACE and FEME (\( p = 0.660 \)) is higher than the one for ACE and FEMF (\( p = 0.1211 \)), the statistical evidence is not enough to segregate the remains with confidence. Secondly, for the comparison between the skeletal parts at the elbow joint, the \( p \)-value obtained indicates a clear acceptance of the null hypothesis for the evaluation of FEMF1 and TIBF (\( p = 0.9729 \)). In contrast, comparison between FEMF1 and TIBE rejects the null hypothesis (\( p = 0.0003 \)) indicating that these elements are unlikely to be associated with each other. Finally, the osteometric comparison of the tali from individuals E and F against the distal tibia of individual E revealed that for both evaluations the null hypothesis was rejected (\( p < 0.05 \)). This suggests that no sorting of these elements is possible by use of osteometric techniques and that other lines of evidence must be evaluated to segregate the remains.
Test 4

Prior osteometric comparison, the $p$-values obtained in both statistical tests (Tables 10 and 11 – Online Resource 1) provide strong evidence which supports that segregation of the skeletal elements based on their shape and size is possible. In both cases, the null hypothesis was accepted only for the associations between SCAPG and HUMG ($p = 0.412$), and HUMH1 and ULH ($p = 0.546$) respectively. Conversely, when opposing individuals where compared the null hypothesis was clearly rejected ($p < 0.05$).

Osteometric sorting model # 3 – Other bone portions

The values for all the osteometric comparisons made to test this model and their respective linear regression values are presented in the Tables 12 and 13 (Test 5) and 14 and 15 (Test 6) of the Online Resource 1.

Test 5

According to the results, there are two possible courses of action for this case. The first one is to segregate HUM(I) from RAD(I), which leads to segregation of HUM(J) and RAD(J). The second option is to segregate HUM(I) from RAD(J), and HUM(J) from RAD(I). If it is considered that the $p$-values obtained from the comparisons between the skeletal elements of the corresponding individuals accepted the null hypothesis ($p > 0.10$), the most adequate decision is to follow the second course of action. This confirms that a correct segregation of the remains is possible.

Test 6

Similar to the previous test, the same two courses of action are available. When assuming the obtained $p$-values, the most reasonable associations are between HUM(K) and FEM(K) ($p = 0.53$ ), and between HUM(L) and FEM(L) ($p = 0.57$). Other possible relationships among these structures will result in rejection of the null hypothesis. The results obtained are correct since they reflex the true association of the bones.
Discussion and conclusions

The analysis of commingled human remains is a complex process that requires a critical and systematic application of objective and reliable techniques for its resolution. Among these, osteometric sorting – defined as “the formal use of size and shape of the skeletal elements for their segregation from one another” [12] - stands as a powerful technique that can be used in conjunction with other sorting methods to maximise the objectivity of the segregation process. Similarly to other aspects of the biological profile, this method strongly relies on osteometric data obtained from samples of known demographic origin, which implies that the accuracy and precision of its results can be diminished when used in individuals derived from biological populations different than the data source [15]. As a result, the reference data used to cement the statistical models need to be representative of the same population as the case specimens in question [4]. For this reason, the main goal of this pilot study was to test the effectiveness of the three osteometric sorting models which were previously developed by Byrd [4] in a sample of modern Colombians. The methods were tested by using an artificially commingled subset of individuals and a reference data sample obtained from this population.

The results obtained in this research, have not only proven that osteometric sorting models can be used and applied in the Colombian population, but also provided a valuable contribution to the general knowledge in a subject poorly explored by forensic anthropology. Furthermore, this study may also have a significant impact for the Colombian legal system, since during the last decades, changes in the court systems around the world have forced forensic sciences to abandon traditional assumptions and move towards a probabilistic approach rooted within Bayesian statistics [18, 19]. As a result human identification methods must be based on solid statistical frameworks for their admissibility in court. Osteometric sorting makes use of statistical tests that are able to provide a measurement of how strong the evidence is against a null hypothesis [4].

Additionally these tests demonstrated that the osteometric sorting models should not be used as a stand-alone technique, but they must be employed together with other independent lines of evidence such as visual pair-matching, elimination, taphonomy, etcetera. As seen in some of the tests results, since the null hypothesis can be either accepted or rejected for both comparisons, the conjoint use with different methods is fundamental to provide strong and accurate results. A careful evaluation and a critical use of all the available methods is pivotal to successfully segregate the remains into their corresponding individuals. The significance levels used during this research were the same for each one of the tests conducted, and only varied between models. However, Byrd [4] recommends that these levels can be larger or smaller, but they need to be assigned according to the particular characteristics and circumstances of the case in question. By considering all these previous options, the re-association of the remains can be much more accurate.

The tests conducted for the purposes of this study revealed that the power of the osteometric sorting method is low and ineffective, when individuals with similar body sizes or similar bone measurements are being
compared. In the test applications where the null hypothesis was either rejected or accepted for both comparisons, the similarity in the bone dimensions appeared to be the main factor that influenced those results. Conversely, when individuals with marked size differences were compared, the tests yielded positive results which demonstrated the power of the method to re-associate skeletal elements of varying dimensions.

As stated by Byrd & Adams [12] and Byrd [4], osteometric models can be created and performed by using only single bone measurements instead of multiple measurements without reducing its overall performance, since both girth and length dimensions tend to have higher correlations with one another. The results of this research prove and augment these findings by showing that the tests developed by using few measurements from length or girth, performed similarly well as the ones were multiple measurements were summed and employed in the models. This fact is particularly important since it demonstrates that osteometric sorting has a great potential to be applied in assemblages were the remains are highly fragmented and deteriorated.

Moreover, Byrd [4] developed a model to compare adjoining skeletal elements by calculating the difference between their areas of articulation. In this study, this model was replicated by using measurements from articulation areas tested in previous studies, such as the hip and knee joints. Additionally, tests were conducted for areas of articulation which have not been previously evaluated in osteometric sorting models, such as the elbow joint and the ankle joint. The results for the tests of the elbow joint (Test 2, osteometric sorting model #2), demonstrated that the evaluation of these areas for purposes of osteometric comparison is viable and can be potentially useful in real casework. In contrast, the results obtained for the comparison of skeletal elements of the ankle joint (distal tibia and talus) suggest that the measurements employed for this test may not be correlated between each other, or that similar to other models, they may be influenced by the similarities of their dimensions. Despite the positive and negative outcomes of all these tests, extensive testing in different and multiple individuals is required to confirm and evaluate the overall effectiveness of these models.

As suggested in the title of this work, the obtained results are part of a pilot research so potential limitations for its application on forensic scenarios may arise. To begin, all the tests developed within this study, only represent small-scale commingled scenarios in which the total number of individuals is a known specific number. Byrd [4] argued that these type of cases present epistemic closure, which occurs when the circumstances of the case allow the investigator to make strong inferences on negative evidence. To exemplify this, assume a case where a left humerus requires being re-associated from a group of two commingled individuals. If one of these individuals has a right humerus which size is significantly larger than the left humerus in question, the most rational option is to segregate this last structure from this individual. The second individual is then, the only one with no evidence against the association made. When evaluating this type of case by means of osteometric models, the assessment of the results is simple and straightforward, since the obtained $p$-value will dictate if the null hypothesis is accepted or rejected. Nevertheless, in a forensic context it is possible to face much more complex scenarios where multiple comparisons must be made in
order to establish the relationships among various skeletal elements, which will result in the generation of multiple $p$-values. A suggested approach to interpret these results was presented by Byrd [4], who developed an omnibus statistic to combine the results of the multiple $p$-values, and eventually obtain a $z$-score for each aggregate of test results, which is later compared against a normal distribution. This statistic can be used to test each of the possible courses of action resulting during the sorting process. This study only tested simple scenarios with relatively simple courses of action, but it did not explore more complex situations like the ones described above. As a result, further research and extensive testing of the osteometric sorting models in more complex samples which involve a larger number of individuals and comparisons is required to solidify the conclusions reached within this work.

Another limitation of this research is that the testing of the statistical models is based on a relatively small sample size. Despite the fact that the results served to confirm and support the findings of previous studies, Byrd & Adams [12] and Byrd [4], emphasize that the statistical models are as good as the reference date they are built upon. This means that the reference data sample must be representative not only for the population of the individuals subjected to analysis, but also of other demographic aspects such as sex, age and ancestry. Although this project made use of a relatively balanced sample in terms of its demographic composition, an increase in its number and the inclusion of individuals from different areas within Colombia and other Latin American countries will be highly beneficial to improve the quality of the statistical models. This also will permit a more representative sample for each one of the skeletal elements considered in this study, and may also allow a wider applicability in different contexts.

Lastly, this project did not explore the effects of sex, age, ancestry, and handedness and secular trends. Despite the empirical results obtained in previous studies suggest that these variables may not have an adverse effect for the development of the models [12], a further study of these issues may allow maximisation of the power of the method, by permitting a refinement of the statistic procedures based on specific components of the population.

In conclusion, based on the results obtained in this research, the osteometric sorting method has proven relevant as a valuable technique for Colombian forensic anthropology to aid to the re-association process of commingled human remains, when used in conjunction with other methods. The osteometric sorting presents multiple advantages for the forensic arena since it is an inexpensive method, it can provide quick and easily interpretable results, it reduces the subjectivity of other segregation techniques, it provides low error rates, it makes use of relatively simple statistical models and it has a considerable power to compare individuals of varying size. Although this pilot study demonstrated some limitations that need to be considered prior its application in a real forensic scenario, further research on these factors will not only provide a solution for those issues, but it will also improve the robustness of the statistical models, it may maximise its credibility and admissibility in court and it will demonstrate the power proffered by this method.
Compliance with ethical standards

Disclosure of possible conflicts of interest

The authors declare that they have no conflicts of interest.

Research involving human participants

Permission for the use and data gathering from the skeletal collection was approved by the Scientific Research Committee of the NILMFS and the Ethics Committee of the University of Dundee. Both institutions granted their permission in accordance with national and international ethical standards.

Acknowledgements

The authors would like to thank to all the staff members of National Institute of Legal Medicine and Forensic Sciences in Bogotá, Colombia which have facilitated and granted the access and use of data from the Skeletal Reference Collection from Colombian Population. Also, the authors thank to the Centre for Anatomy and Human Identification of the University of Dundee for their valuable help while conducting this research.

References


APPENDIX 1

List of sources and names of variables used in this study

<table>
<thead>
<tr>
<th>SKELETAL ELEMENT</th>
<th>VARIABLES</th>
</tr>
</thead>
</table>
| **SCAPULA**      | Height 
(Anatomical breadth) 
(SCH) [13] |
|                  | Breadth 
(Anatomical length) 
(SCBR) [13] |
|                  | Maximum height of the glenoid 
(MxGLH) [14] |
|                  | Maximum breadth of the glenoid 
(MxGLB) [14] |
|                  | Maximum length 
(HUMxL) [13] |
|                  | Vertical diameter of head 
(VDH) [13] |
|                  | Anterior-posterior diameter of head 
(APDH) [14] |
| **HUMERUS**      | Maximum diameter at midshaft 
(HMxDM) [13] |
|                  | Minimum diameter at midshaft 
(HMnDM) [13] |
|                  | Epicondylar breadth 
(HEB) [13] |
|                  | Capitulum Trochlea Breadth 
(CTBR) [14] |
| **RADIUS**       | Maximum length 
(MLRAD) [13] |
|                  | Sagittal diameter at midshaft 
(RAPD) [13] |
|                  | Transverse diameter at midshaft 
(RTRD) [13] |
|                  | Maximum shaft diameter on the radial tuberosity 
(RMxDT) [4] |
|                  | Maximum diameter of radial head 
(MDRH) [4] |
| **ULNA**         | Maximum length 
(UML) [13] |
|                  | Physiological length 
(UPL) [13] |
|                  | Semilunar breadth 
(SLB) [12] |
|                  | Dorso-Volar diameter 
(UDVD) [13] |
<table>
<thead>
<tr>
<th>Skeletal Element</th>
<th>Variables recorded</th>
</tr>
</thead>
</table>
| OS COXAE         | Transverse diameter ($UTRD$) [13]  
|                  | Minimum circumference ($UMC$) [13]  
|                  | Acetabular height ($ACH$) [15]  
|                  | Height ($OCH$) [13]  
|                  | Iliac breadth ($IBR$) [13]  
| FEMUR            | Bicondylar length ($FBIL$) [13]  
|                  | Epicondylar breadth ($FEB$) [13]  
|                  | Maximum head diameter ($FHMD$) [13]  
|                  | Anterior-Posterior subtrochanteric diameter ($APSTD$) [13]  
|                  | Medial-Lateral subtrochanteric diameter ($MLSTD$) [13]  
|                  | Midshaft circumference ($FMC$) [13]  
| TIBIA            | Length ($TLN$) [13]  
|                  | Maximum proximal epiphyseal breadth ($TMPEB$) [13]  
|                  | Maximum distal epiphyseal breadth ($TMDEB$) [13]  
|                  | Maximum diameter at the nutrient foramen ($TMDNF$) [13]  
|                  | Transverse diameter at the nutrient foramen ($TTDNF$) [13]  
| FIBULA           | Maximum length ($FML$) [13]  
|                  | Maximum diameter at midshaft ($FDM$) [13]  
| TALUS            | Trochlear breadth ($TRB$) [16]  

**Table 1.** List of variables recorded for each skeletal element.
APPENDIX 2

List of measurement values of the skeletal elements used for the comparisons in this study

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>HUMxL</th>
<th>VDH</th>
<th>APDH</th>
<th>HMxDM</th>
<th>HEB</th>
<th>MLRAD</th>
<th>RAPD</th>
<th>RTRD</th>
<th>UML</th>
<th>UPL</th>
<th>UDV</th>
<th>UTRD</th>
<th>UMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMA</td>
<td>341</td>
<td>48,2</td>
<td>44,94</td>
<td>26,08</td>
<td>64</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>HUMB</td>
<td>284</td>
<td>4,27</td>
<td>39,8</td>
<td>21,38</td>
<td>59</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>HUMA1</td>
<td>337</td>
<td>50,18</td>
<td>45,22</td>
<td>24,42</td>
<td>63</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>RADA</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>257</td>
<td>13,94</td>
<td>16,01</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>RADA1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>258</td>
<td>13,6</td>
<td>15,55</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>RADB</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>221</td>
<td>12,46</td>
<td>15,68</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>ULA</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>274</td>
<td>248</td>
<td>16,03</td>
</tr>
<tr>
<td>ULB</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>241</td>
<td>212</td>
<td>14,17</td>
</tr>
<tr>
<td>ULB1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>240</td>
<td>210</td>
<td>14,87</td>
</tr>
</tbody>
</table>

Table 1. Measurement values for the skeletal elements evaluated in Test 1.

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>FBIL</th>
<th>FEB</th>
<th>FHMD</th>
<th>APSTD</th>
<th>MLSTD</th>
<th>FMC</th>
<th>TLN</th>
<th>TMDNF</th>
<th>TTDNF</th>
<th>FML</th>
<th>FMDM</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEMD1</td>
<td>410</td>
<td>76</td>
<td>38,71</td>
<td>22,67</td>
<td>28,99</td>
<td>71</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>FEMC</td>
<td>400</td>
<td>68</td>
<td>39,73</td>
<td>26,09</td>
<td>30,9</td>
<td>78</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>FEMD</td>
<td>409</td>
<td>76</td>
<td>39,69</td>
<td>23,21</td>
<td>27,78</td>
<td>70</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>TIBD1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>323</td>
<td>30,04</td>
<td>19,74</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>TIBC</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>325</td>
<td>31,83</td>
<td>23,56</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>TIBD</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>323</td>
<td>29,7</td>
<td>20,53</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>FIBD1</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>318</td>
<td>13,57</td>
<td></td>
</tr>
<tr>
<td>FIBD</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>318</td>
<td>14,05</td>
<td></td>
</tr>
<tr>
<td>FIBC</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>318</td>
<td>14,72</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Measurement values for the skeletal elements evaluated in Test 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>ACE</th>
<th>FEME</th>
<th>FEMF</th>
<th>FEMF1</th>
<th>TIBE</th>
<th>TIBF</th>
<th>TIBE1</th>
<th>TALE</th>
<th>TALF</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACH</td>
<td>43,36</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>FJMD</td>
<td>--</td>
<td>38,97</td>
<td>45,37</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>FEB</td>
<td>--</td>
<td>--</td>
<td>79</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>TMPEB</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>67</td>
<td>74</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>TMDNF</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>31,23</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>TRB</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>26,57</td>
<td>32,89</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Measurement values for the skeletal elements of the hip joint evaluated in Test 3.


<table>
<thead>
<tr>
<th>Variable</th>
<th>SCAPG</th>
<th>HUMG</th>
<th>HUMH</th>
<th>HUMH1</th>
<th>ULH</th>
<th>ULG</th>
</tr>
</thead>
<tbody>
<tr>
<td>MxGLH</td>
<td>42,38</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>MxGLB</td>
<td>28,35</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>VDH</td>
<td>--</td>
<td>50,18</td>
<td>39,09</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>APDH</td>
<td>--</td>
<td>45,22</td>
<td>35,48</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>CTBR</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>35,92</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>SLB</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>17,5</td>
<td>24,14</td>
</tr>
</tbody>
</table>

Table 4. Measurement values for the skeletal elements of the shoulder joint evaluated in Test 4.

<table>
<thead>
<tr>
<th>Variable</th>
<th>HUM(I)</th>
<th>RAD(I)</th>
<th>HUM(J)</th>
<th>RAD (J)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMxDM</td>
<td>20,03</td>
<td>--</td>
<td>18,1</td>
<td>--</td>
</tr>
<tr>
<td>HMnDM</td>
<td>14,87</td>
<td>--</td>
<td>13,11</td>
<td>--</td>
</tr>
<tr>
<td>RMxDT</td>
<td>--</td>
<td>14,87</td>
<td>--</td>
<td>13,01</td>
</tr>
<tr>
<td>MDRH</td>
<td>--</td>
<td>22,97</td>
<td>--</td>
<td>19,75</td>
</tr>
</tbody>
</table>

Table 5. Measurement values from the humeri and radii, evaluated in Test 5.

<table>
<thead>
<tr>
<th>Variable</th>
<th>HUM(K)</th>
<th>FEM(K)</th>
<th>HUM(L)</th>
<th>FEM (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMxL</td>
<td>330</td>
<td>--</td>
<td>284</td>
<td>--</td>
</tr>
<tr>
<td>FBIL</td>
<td>--</td>
<td>480</td>
<td>--</td>
<td>405</td>
</tr>
</tbody>
</table>

Table 6. Measurement values from the humeri and femora, evaluated in Test 6.