Article

A Non-Racial Approach to Assessing Group Membership of Victims in a Mass Grave Using Cranial Data

John Albanese * and Alyssa Di Iorio

Department of Integrative Biology, University of Windsor, Windsor, ON N9B 3P4, Canada
* Correspondence: albanese@uwindsor.ca; Tel.: +1-(519)-253-3000 (ext. 3973); Fax: +1-(519)-971-3621

Abstract: In some jurisdictions, race, ancestry, or population affinity have been used for historical and po-litical, rather than biological, reasons in forensic anthropology when identifying individuals. The approach persists even though the genetic and skeletal data clearly demonstrate that human variation does not cluster into these groups. For over 60 years, these methods have consistently performed poorly when independently tested using large samples. By racializing the deceased, these methods have further marginalized the living. However, there is a need in the investigation of genocide and human rights violations to demonstrate that a specific group was targeted. Without relying on the outdated typological concepts of human variation, in this paper we present preliminary results for a method that can be used in a mass grave context to demonstrate that a specific group was targeted. Using samples from two identified reference collections, we created subsamples from one relatively homogeneous collection to model various mass grave scenarios and used the relatively heterogenous sample from the other collection as a reference for com-parison. In scenarios that varied by sample size and sex, it was possible to determine that a specific group was targeted if the sample size in a hypothetical mass grave was greater than 25 for a multi-sex sample, when sex is not known, and a minimum of 14 if sex could be estimated.

Keywords: forensic anthropology; cranial measurements; critique of race; ancestry; racialization; human rights; genocide

1. Introduction

Forensic anthropologists can assist with the identification of an unknown individual by constructing a biological profile by estimating sex, stature, and age at death from skeletal remains. In some jurisdictions, ancestry (sometimes referred to as race or population affinity) is also considered part of identification for historical and political reasons rather than biological reasons [1]. There are many theoretical, methodological, and ethical problems with the continued use of these typological approaches in forensic anthropology. Theoretically, osteometric and genetic variation does not cluster in racial groups, continental origins, and national affinities—for a systemic review in forensic anthropology see [1,2]. Methodologically, when tested systematically using independent samples, methods perform very poorly when estimating group membership, matching premortem documents in only 36–50% of cases [1,3]. These poor results are consistent with some of the earliest tests of these methods beginning in the 1960s, for example, [4]. Changes in reference samples, alternative statistical approaches, new terminology, software updates, etc., have not resulted in any improvements in providing information that could be useful in a forensic investigation [1]. Additionally, linking sex estimation and stature estimation to group membership can make these methods more difficult to apply, while compromising accuracy and utility—see [1,5]. Ethically, there are at least two major problems with these typological approaches. First, forensic anthropologists should be using the best methods that are available and stop using methods that provide wrong information in over 50% of cases [1,3]. Second, the continued racialization of the dead only serves to reinforce

existing power structures that marginalize the living [6]. With a privileged place at the intersection of science and law, forensic anthropologists can have an enormous influence on how individuals and groups are racialized and marginalized by what they say and how they say it [7].

For individual cases, assessing ancestry/race/affinity should be avoided because methods provide wrong information that would seriously compromise an investigation in the majority of cases. Forensic anthropologists should adopt a more robust protocol for identification that does not include estimating race, ancestry, or affinity [1,3,6]. However, there is a need to prove that a specific group was targeted in investigations of genocide and crimes against humanity. In this paper, we present an approach to estimate group membership that can be applied to mass graves. Although mass graves are defined differently by various researchers, based on the number of individuals required for a burial site to be considered a mass grave, most researchers agree that a burial site must contain remains that are tightly packed and indiscriminately placed to be considered a mass grave [8]. In other words, the grave is an attempt to hide the evidence of mass murder and deny the dead the locally appropriate mortuary rituals and practices and is thus a continuation of structural violence inflicted on the dead and the survivors.

Without relying on the outdated typological concepts of human variation, we present some preliminary results for an approach that can be used within the context of a mass grave to demonstrate that a specific group was the target of violence. In this research we use carefully selected research samples from two different identified reference collections (IRCs). Subsamples from the relatively homogeneous Coimbra collection are used to model various mass grave scenarios involving one or multiple sexes and a range of sample sizes. A sample from the relatively heterogenous Terry collection is used as a reference sample for comparison.

2. Materials and Methods

2.1. The Terry Collection

The Terry collection was amassed by Robert J. Terry and Mildred Trottier over the first six decades of the 20th century—for details see [9]. They collected skeletons from cadavers used for anatomical instruction at the medical school at Washington University in St. Louis, Missouri. The cadavers originated primarily from unclaimed individuals who died at various hospitals and institutions from the surrounding area; while a subset of the cadavers included individuals who had bequeathed their remains to be used for anatomical instruction when it became legally possible and socially acceptable to do so. All IRCs are biased in some way due to the collections process and other factors, but bias is a problem if it is ignored in research design and bias can be exploited to address specific research questions [10]. The Terry collection can be considered relatively heterogenous because the collection includes individuals with a wide range of ages at death and years of birth and was drawn from a wide range of geopolitical locations over a long period covering a great deal of political and economic change. For example, some individuals in the collection suffered from the impacts of poverty and economic disparity during their entire lives, while others who died during the Great Depression may have lived in extreme poverty only at the end of their lives—for details about these skeletal manifestations see [2,7]. The study sample used in this research is a subset of the sample previously used to develop both sex and stature estimation methods [5,11]. The sample was carefully selected using year of birth together with age at death to sample a wide range of human variation in *Homo sapiens* associated with secular changes and mortality bias [11]. For the current research, the sample from the Terry collection was used as a relatively heterogenous reference sample.

2.2. The Coimbra Collection

The Coimbra collection (*Colecção de Esqueletos Identificados*), which is curated at the University of Coimbra, includes the identified remains of individuals from the main cemetery in the City of Coimbra, Portugal, who died between 1904 and 1936 [12,13]. The

collection is relatively homogeneous due to the collection process and the source of the remains. Most of the individuals were included in the collection in two narrow phases of collection from the same cemetery. Most were born and lived in the District of Coimbra, or adjacent districts, and they were from the same socio-economic and political classes. They lived, died, and were buried and exhumed under similar social, political, economic, and legal circumstances. The collection is derived from a relatively homogeneous local community similar to a group that could be targeted in acts of genocide. For this research, the sample from the Coimbra collection was used to create various mass grave scenarios that varied by sample size and sex.

2.3. Cranial Measurements

We pursued a practical approach and used 10 standard cranial measurements that capture overall cranial size and shape and that are relatively resistant to premortem changes such as tooth loss, perimortem changes and trauma, and taphonomic processes including burial and recovery. Table 1 includes a list of measurements and a description of each measurement used in this research [14]. These measurements were also selected because they can be easily collected using spreading and sliding calipers with a minimal amount of training. All measurements were collected by one researcher (JA).

Table 1. Descriptions of cranial measurements [14].

Measurements	Definition
Maximum cranial length	Distance of glabella from opisthocranion in the mid sagittal plane measured in a straight line.
Maximum cranial breadth	Maximum width of the skull perpendicular to the mid-sagittal plane, wherever it is located, with the exception of the inferior temporal line and the immediate area surround the latter.
Bizygomatic breadth	Direct distance between both zygia located at their most lateral points of the zygomatic arches.
Cranial height	Direct distance from the lowest point on the anterior margin of the foramen magnum, basion, to bregma.
Cranial base length	Direct distance from nasion to basion.
Basion prosthion	Direct distance from basion to prosthion.
Maximum alveolar breadth	Maximum breadth across the alveolar borders of the maxilla, measured on the lateral surfaces at the location of the second maxillary molars.
Upper facial height	Direct distance from nasion to prosthion.
Minimum frontal breadth	Direct distance between the two frontotemporale.
Mastoid length	Projection of the mastoid process below, and perpendicular to, the eye–ear plane in the vertical plane.

Many of the crania in the Terry collection were sectioned when used for anatomical instruction [9]. Measurements were not taken across the cuts to the cranium because the amount of bone removed by the saw could not be estimated. The total sample from the Terry collection of 111 cases included 48 individuals whose sex was documented as female, and 63 individuals whose sex was documented as male. The sample size for the Coimbra collection varied with each trial to recreate various mass grave scenarios with different numbers of overall cases for all sexes, and males and females separately. Sample size for each of the 60 scenarios are presented in Tables 2–4.

Table 2. Allocation accuracies using logistic regression with decreasing sample sizes from the Coimbra collection used to model mass grave scenarios with all sexes. Sample size from the Terry collection was held constant (n = 111) as a reference sample.

	Coimbra Collection			Terry Collection (n = 111)		
n	**Coimbra**	**Terry**	**% Correct**	**Coimbra**	**Terry**	**% Correct**
127	107	20	84.3	18	93	83.8
100	82	18	82.0	15	96	86.5
81	61	20	75.3	14	97	87.4
73	54	19	74.0	11	100	90.1
66	48	18	72.7	12	99	89.2
52	39	13	75.0	8	103	92.8
51	38	13	74.5	9	102	91.9
49	36	13	73.5	9	102	91.9
46	28	18	60.9	10	101	91.0
38	27	11	71.1	4	107	96.4
34	21	13	61.8	6	105	94.6
28	18	10	64.3	6	105	94.6
26	14	12	53.8	6	105	94.6
24	10	14	41.7	7	104	93.7
23	10	13	43.5	6	105	94.6
21	11	10	52.4	6	105	94.6
15	7	8	43.8	4	107	96.4
11	2	9	18.2	1	110	99.1
8	3	5	37.5	1	110	99.1
7	1	6	14.3	0	111	100

Table 3. Allocation accuracies using logistic regression with decreasing sample sizes from the Coimbra collection used to model mass grave scenarios with individuals whose sex was documented as female. Sample size from the Terry collection was held constant (n = 48) as a reference sample.

	Coimbra			Terry (n = 48)		
n	**Coimbra**	**Terry**	**% Correct**	**Coimbra**	**Terry**	**% Correct**
64	59	5	92.2	8	40	83.3
58	53	5	91.4	8	40	83.3
50	44	6	88.0	8	40	83.3
48	41	7	85.4	7	41	85.4
46	41	5	89.1	7	41	85.4
41	35	6	85.4	7	41	85.4
37	29	8	78.4	6	42	87.5
36	30	6	83.3	7	41	85.4
31	26	5	83.9	7	41	85.4
29	22	7	75.9	7	41	85.4
26	20	6	76.9	7	41	85.4

Table 3. *Cont.*

	Coimbra			Terry (n = 48)		
n	**Coimbra**	**Terry**	**% Correct**	**Coimbra**	**Terry**	**% Correct**
25	19	6	76.0	5	43	89.6
23	15	8	65.2	5	43	89.6
21	15	6	71.4	5	43	89.6
16	10	6	62.5	6	42	87.5
14	8	6	57.1	4	44	91.7
13	5	8	38.5	5	43	89.6
11	7	4	63.6	3	45	93.8
7	2	5	28.6	1	47	97.9
6	6	0	100	0	48	100

Table 4. Allocation accuracies using logistic regression with decreasing sample sizes from the Coimbra collection used to model mass grave scenarios with individuals whose sex was documented as male. Sample size from the Terry collection was held constant (n = 63) as a reference sample.

	Coimbra			Terry (n = 63)		
N	**Coimbra**	**Terry**	**% Correct**	**Coimbra**	**Terry**	**% Correct**
63	51	12	81.0	11	52	82.5
56	45	11	80.4	10	53	84.1
50	39	11	78.0	7	56	88.9
49	39	10	79.6	9	54	85.7
47	37	10	78.7	9	54	85.7
45	33	12	73.3	9	54	85.7
40	28	12	70.0	9	54	85.7
37	28	9	75.7	3	60	95.2
34	25	9	73.5	4	59	93.7
27	19	8	70.4	5	58	92.1
24	18	6	75.0	1	60	95.2
23	17	6	73.9	3	60	95.2
19	12	7	63.2	2	61	96.8
15	10	5	66.7	1	62	98.4
14	7	7	50.0	2	61	96.8
12	6	6	50.0	1	62	98.4
11	4	5	44.4	3	60	95.2
9	5	4	55.6	1	62	98.4
6	3	3	50.0	1	62	98.4
4	4	0	100	0	63	100

2.4. Statistical Approach

Logistic regression is a robust probabilistic approach for predicting a dichotomous outcome that has occasionally been used in Forensic Anthropology for sex estimation—for early examples see [11,15]. In this research, this statistical approach is used to compare the mass grave subsamples from the Coimbra collection to the reference sample from the

Terry collection. Essentially, logistic regression is used to allocate individuals to one of two groups: Coimbra collection subsamples ("mass grave") or not Coimbra collection subsamples ("not mass grave"). Given this approach, an allocation accuracy of greater than 50% in allocating individuals from the Coimbra subsamples to the Coimbra subsamples instead of the Terry collection sample suggests that a majority of the individuals in the hypothetical mass grave are more similar to each other than to the more diverse reference sample from the Terry collection.

Logistic regression calculates a *p*-value between zero and one for each case. The default setting is to have 0.5 as the sectioning point. This value can be adjusted to maximize allocation accuracy, but, in this research, it was kept at 0.5 for consistency across trials. Given the coding of the data, a *p*-value greater than 0.5 allocated an unknown to the Terry collection sample, and a *p*-value less than 0.5 allocated an unknown to the Coimbra collection subsamples. The *p*-value is also the probability of a correct assessment. For example, a *p*-value of 0.9 would allocate the case to the Terry collection sample and can be interpreted as a 90% probability that the individual is from the Terry collection sample. Conversely a *p*-value of 0.1 would allocate the case to the Coimbra subsamples with a probability of 90% ($1 - 0.1 = 0.9$). Generally, one of the strengths of logistic regression is that the *p*-value is used as an indicator of certainty in any one case independent of the overall allocation accuracy. However, in this research the probability of any one correct allocation is less of an issue. What matters most is the pattern of *p*-values of the subsamples modelled as a mass grave. The *p*-values for each case were assessed in more detail for the all-sexes scenarios that provided the best results with the largest subsamples, and with the smallest subsamples (n = 26) that met the minimum parameters for useability.

### 2.5. Selection of Subsamples to Model Various Mass Grave Scenerios

Various scenarios were considered for all-sexes subsamples and sex-specific subsamples. The first step was to run the analysis with the entire Coimbra all-sexes sample. The Coimbra sample was then randomly resampled with decreasing sample sizes using complex random sample selection feature in SPSS 28, wherein the Terry collection sample remained the same and a specified percentage of the Coimbra collection sample was randomly sampled. First, a random sample of 80% of the entire Coimbra collection sample was selected as a model for a mass grave, then 60% was randomly sampled, then 40% was randomly sampled, etc. After these initial major steps of 20% intervals in sampling, the Coimbra collection was resampled to fill gaps in sample size, with a focus on smaller sample sizes to test the limits of applicability of the method. Twenty trials of varying sample sizes (n = 7 to 127) were analyzed using logistic regression. The resampling was stopped at 20 trials because a clear pattern began to emerge with the results. The approach was repeated for those individuals wherein sex was documented as female (n = 6 to 64), and again for those whose sex was documented as male (n = 4 to 63).

This approach for an all-sexes, and sex-specific analysis was pursued to assess the parameters of applicability of the method. In some cases, certain sexes or genders are specifically targeted for violence and are included in a mass grave. In other cases, all sexes and genders are targeted. Biological sex can be defined using various criteria, but using relatively straight forward chromosomal criteria, there are at least five sexes that occur in humans at measurable frequencies: XO, XX, XY, XXY, XYY—for a more comprehensive and nuanced review see [16]. One goal of this research is to assess if there are any benefits of using a sex-specific approach that relies on a binary conceptualization of sex. Does controlling for one of the two more common biological sexes (XX or XY) and/or the more commonly documented sexes, female or male, provide better results? Another goal is to assess the parameters of applicability of the method when a binary approach to sex estimation is not applicable, sex cannot be reliably estimated, or marginalized genders or sexes (not XX or XY) have been disproportionately targeted.

3. Results

An allocation accuracy of greater than 50% for a subsample from the Coimbra collection indicates that the individuals included in the hypothetical mass grave have more in common with each other than with the reference sample from the Terry collection. This threshold of 50% was exceeded for the all-sexes scenarios when the sample size for the mass grave was greater than 25 cases (Table 2). The 50% threshold was met with smaller sample sizes for sex-specific trials. For scenarios that included individuals whose sex was documented as female (Table 3), allocation accuracy for the Coimbra collection mass grave subsamples consistently exceeded the 50% threshold when the sample size was 14 individuals or higher. When the sample was below 14 individuals, the accuracy varied widely from 28.6% (n = 7), to 38.5% (n = 13), to 100% (n = 6). For scenarios that included individuals whose sex was documented as male (Table 4), the Coimbra collection mass grave subsamples consistently met or exceeded the 50% threshold with a sample size of about 12 to 14 individuals. As with the females, the allocation accuracy below this sample size produced both the lowest allocation accuracy of 44.4% (n = 11) and the highest allocation accuracy of 100% (n = 4). It is worth stating explicitly that the 100% accuracy for the smallest sex-specific models are likely due to chance since the results are not consistent with other results wherein accuracy is correlated to sample size.

Excluding the smallest sample sizes (n = 4 males and n = 6 females), which had the highest allocation accuracies (100%), there is a clear pattern wherein allocation accuracy increases with an increase in size of the mass grave sample. This pattern is clearly visible in Figure 1, a scatter plot of sample size by allocation accuracy. The other pattern that is clear in Figure 1 is that, for any given sample size, the allocation accuracy for the sex-specific scenarios always exceeds the all-sexes scenarios.
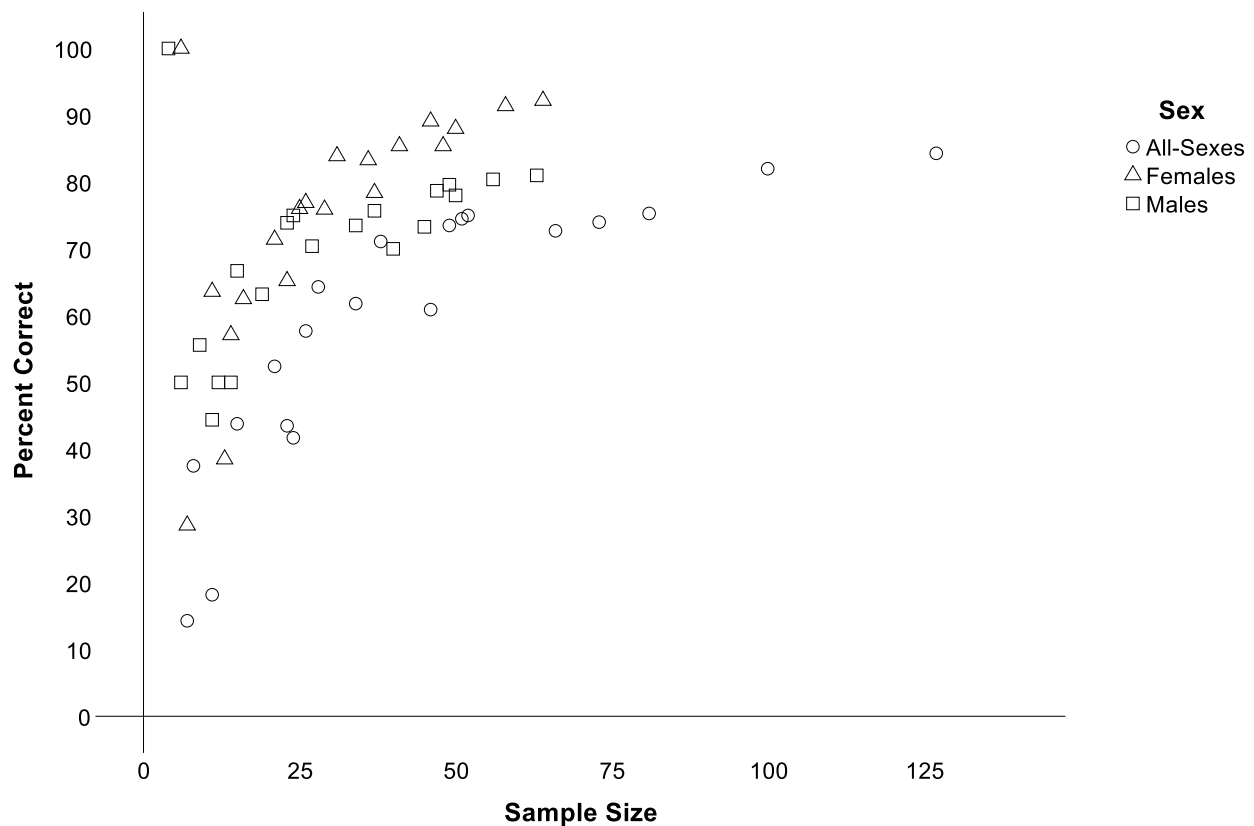

Figure 1. A scatter plot of sample size by allocation accuracy for the Coimbra collection subsamples for 60 trials modeling various mass grave scenarios.

The *p*-values for each case were reviewed in detail for two specific trials for the all-sexes scenarios to assess the pattern of allocation. The *p*-values for the scenario with the

best results using the entire Coimbra collection sample are presented graphically in Figure 2. The *p*-values for the subsamples (n = 26) that met the minimum parameter (50% allocation accuracy) for useability are presented in Figure 3. For both figures, the first column includes *p*-values between 0.0 and 0.09, the second column includes *p*-values between 0.1 and 0.19, the third column includes *p*-values between 0.2 and 0.29, etc. The white portion of each column represents the Coimbra collection individuals, and the shaded portion of each column represents the Terry collection individuals. Any given column may include individuals from either collection. For example, the first column in Figure 2 includes four individuals from the Terry collection sample allocated to the Coimbra collection sample, and over 50 individuals from the Coimbra collection sample allocated to the Coimbra collection sample. The pattern in Figure 2 clearly demonstrates that the majority of the sample from the Coimbra collection was allocated to the Coimbra collection with *p*-values from 0.0 to 0.29 (probabilities of 99% to 71%, respectively), as seen in the white portions of first three columns. The strong *p*-values from each individual case confirm the relative homogeneity of the Coimbra collection sample. The results in Figure 3 appear to be different because of the much smaller sample size for the Coimbra collection subsamples. However, these results are consistent with Figure 2. For this scenario wherein the Coimbra collection subsamples included 26 individuals, 14 of those individuals (53.8%) were allocated to the Coimbra collection. Of those 14 cases, 13 cases out of 26 (i.e., 50%) had *p*-values between 0.0 and 0.29 (probabilities of 99% to 71%, respectively). Even after excluding a case from the Coimbra collection with an ambiguous *p*-value of 0.44 (the one case in white in the column 0.4 in Figure 3), the overall allocation accuracy is still 50%.
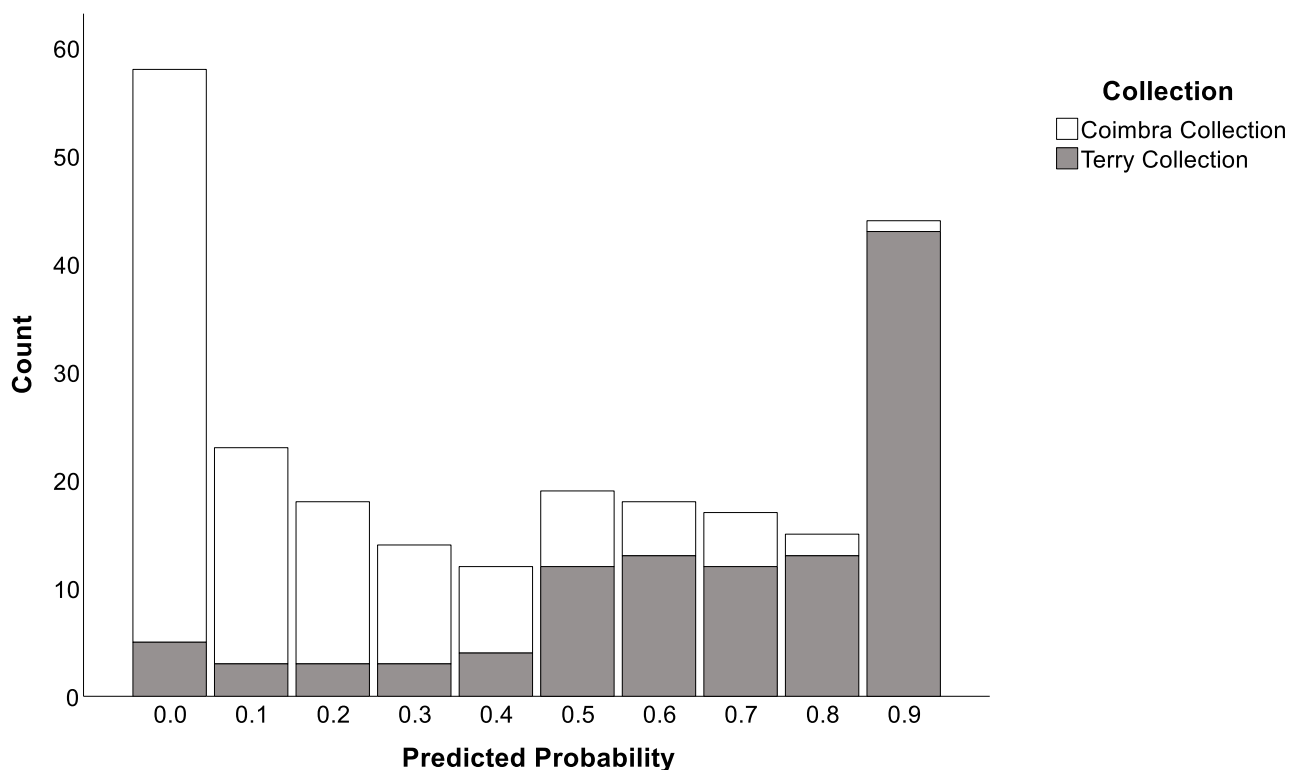
Figure 2. Distribution of *p*-values for the all-sexes Terry collection sample (n = 111) and the entire all-sexes Coimbra collection sample (n = 127).
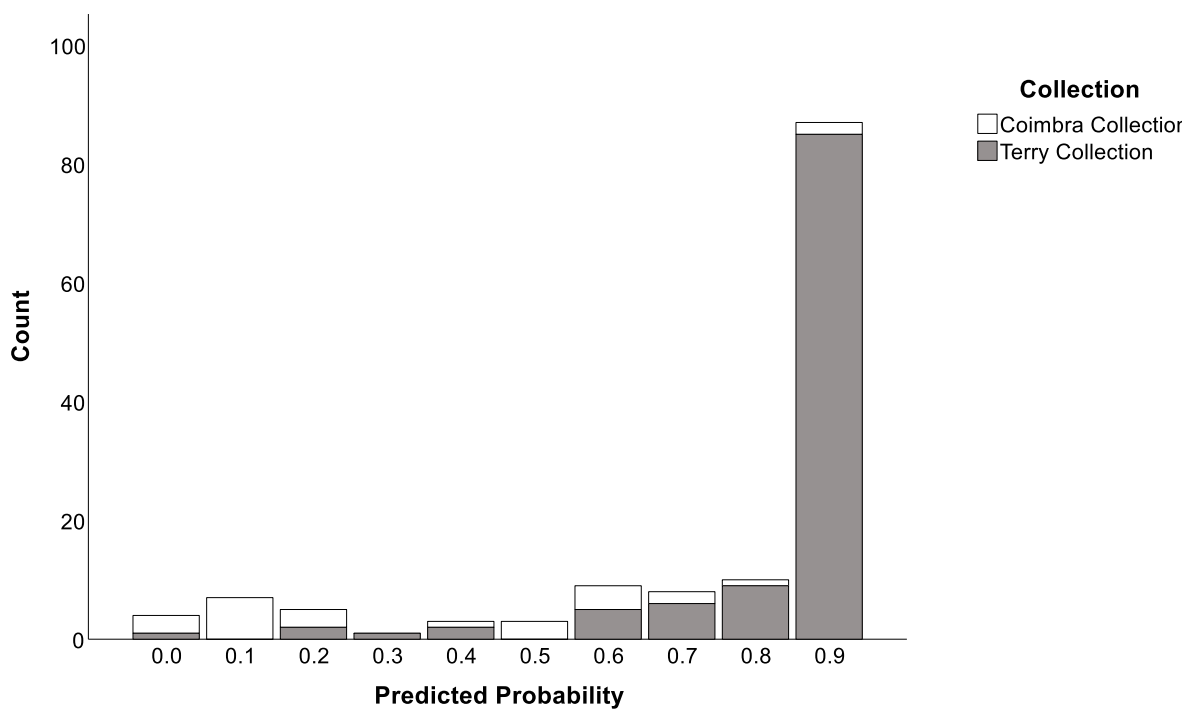
Figure 3. Distribution of *p*-values for the all-sexes Terry collection sample (n = 111) and the smallest all-sexes Coimbra collection subsamples (n = 26) that met the minimum 50% allocation accuracy criterion.

4. Discussion

Allocation accuracies varied with subsample sizes and sex-specific approaches in various scenarios, but consistently demonstrated the utility of the method for assessing that a specific relatively homogeneous group was targeted and included in a mass grave. Allocation accuracies of greater than 50% for the various subsamples modelled as mass grave scenarios were possible, provided that minimum samples (n > 13 for sex-specific and n > 25 for all-sexes) were available for analysis. The modeling of the Coimbra collection subsamples demonstrates that it is possible to assess that those individuals who were murdered and included in the mass grave have more in common with each other than with an external, independent reference sample.

Controlling for some variation related to commonly documented sexes (male and female) consistently provided better results and worked well with smaller samples. These results are statistically expected since a sex-specific subsample, regardless of how sex is defined, will include a narrower range of variation for any given subsample. A male-only sample will exclude the smallest individuals from a subsample, and the female-only sample will exclude the largest individuals from the subsample. Either approach will truncate the range of variation. Controlling for some of the variation associated with documented sex necessarily made the Coimbra subsamples more homogeneous and allowed for good results (over 50%) with smaller sample sizes. Although the all-sexes approach requires a larger sample, it is a more robust approach for several reasons. First, sex cannot always be reliably estimated. It can be difficult to assess which individuals might be larger females versus smaller males. In other cases, individuals who identify as men (gender) and are targeted because of their gender may have a skeletal phenotype that is consistent with a "female pattern" or vice versa. Although counter-intuitive, a sex-specific approach based on various biological assumptions may perform worse than an all-sexes approach when a specific gender was targeted because biological sex does not neatly map onto gender. Additionally, the all-sexes approaches can be pursued without linking this method to sex estimation. Individual sex and group membership remain two separate independent questions, and errors in sex estimation will not have an impact on the assessment of group membership—for a well-documented case involving stature estimation see [5]. Overall,

the all-sexes approach should be used when a binary approach to sex estimation is not applicable, sex cannot be reliably estimated, or marginalized genders or sexes (not XX or XY) have been disproportionately targeted.

An added benefit of this approach that focuses on the entire group of victims rather than any one individual is that missing data for any one case (for example, due to severe cranial trauma), or the ambiguous *p*-value in any one case (see Figure 3), will not compromise the utility of the method. The measurements collected for most individuals in the mass grave can be effectively used to demonstrate that a specific relatively homogeneous group was the target of violence, and a clear cluster of strong *p*-values can be used to support the conclusion (see Figures 2 and 3). The results wherein allocation accuracy exceeded 50% were consistent when sample sizes fluctuated slightly. For example, when the subsamples for the all-sexes scenario (Table 2) ranged from 49 to 52 individuals, the allocation accuracy ranged from 73.5% to 75%. All the results in this accuracy range strongly demonstrate the homogeneity of the subsample used to model a mass grave. Having to exclude a few cases because of missing data has minimal impact on the utility of the method provided the sample size minimums are met.

Additional research is needed to establish the full parameters of applicability and utility of the method using different predictor variables. Logistic regression allows for a great deal of flexibility in which variables can be used as predictor variables because the data do not need to be normally distributed. It is possible to use different measurements and possibly fewer measurements, and to combine measurements with morphological data and dental non-metric traits as predictor variables. However, future research should be conducted to confirm the minimum number of variables required, and which combinations can be used effectively. In this research we used 10 standard cranial measurements that capture overall cranial size and shape. These standard measurements are often already being collected as part of a standard protocol in many investigations. Thus, the approach described here can be applied to data collected from individuals from mass graves that have already been excavated. The method has potential beyond the investigation of genocide and human rights violations. The method can be applied in some bioarchaeological contexts as well. This research was initiated, in-part, by a need to establish the parameters of applicability of the method to assess the relationships of the deceased included in a Late Bronze Age monumental tomb on the island of Kefalonia, Greece, without relying on destructive methods —for detailssee [17].

Skeletal analysis is one of two critical steps to determine that a specific group was targeted and included in a specific grave. After the relative homogeneity of the victims in a mass grave is confirmed with this new method, the ethnicity of the targeted group must be assessed using material cultural objects found in the mass grave with the victims. To be clear, we are not attempting to estimate "ethnicity" as a euphemism for race, ancestry, or affinity. The skeletal data are not used to estimate "ethnicity." The true ethnicity of a group is assessed with the material cultural objects and personal belongings of those murdered and placed in a mass grave. The skeletal data would only indicate that the victims were a relatively homogeneous group when compared to an external reference sample. A similar approach using material cultural objects can be used to assess if certain genders were disproportionately targeted.

In addition to providing evidence supporting the utility of a new method, the results from this research also provide additional evidence against using a typological approach in forensic anthropology. The Coimbra collection subsamples consisting of European-born individuals did not cluster with those individuals from the Terry collection who were described as "White" in various documents—for details about documentary data see [9]. Furthermore, those individuals who were described in various documents as "White" and "Black" from the Terry collection were consistently grouped together. The results in Tables 2–4 indicate that the allocation accuracy for the Terry collection sample starts high at 82.5% and increases to 100% as the Coimbra collection subsamples decrease. However, a more rigorous approach is to hold sample sizes from each collection approximately equal.

For the all-sexes comparison (Table 2), the Terry collection sample included 111 cases, and the allocation accuracy for this reference sample was 83.8%. Comparable samples from the Coimbra collection are 100 and 127 individuals, and allocation accuracy for these Coimbra samples were 82.0% and 84.3%, respectively. Similarly, for the female-specific scenario wherein sample size was the same for each collection (n = 48), the allocation accuracy was 83.3% for the Terry collection individuals and 85.4% for the Coimbra collection subsamples (Table 3). Lastly, for the male-specific scenario the maximum sample size was the same at 63 individuals for the Coimbra collection sample and the Terry collection, and allocation accuracy was 82.5% for the Terry collection sample and 81.0% for the Coimbra collection sample. It is possible to predict collection membership with an allocation accuracy exceeding 80% with minimal bias in accuracy by collection. The historical, socioeconomic, and political context of the collection has much more explanatory power than race/ancestry/affinity when investigating patterns of human variation (2,5,10,11,13,18).

The results from this research demonstrate that the method for assessing group homogeneity can work in a mass grave context, and that approaches that use race, ancestry, or affinity are not effective at capturing patterns of human variation and predicting group membership. However, the results do not show that the variation in the subsamples from the Coimbra collection can, or should, be extrapolated nationally, and the results do not demonstrate a pattern of "population affinity" for Portugal—see [10]. The homogeneity of the Coimbra collection subsamples is due to the samples being derived from the mortality sample of a group from a very specific time, place, and political and socioeconomic context. The method works because it assesses temporally, geographically, and economically specific biocultural variation for a specific local group. A different pattern of variation is necessarily expected with living conditions that vary through time and space. For example, in the context of assessing patterns of secular change, the mean maximum femur length for males from the Coimbra collection born in the 19th century was not significantly different from the mean for females from the Terry collection when controlling for year of birth [18]. When looking at secular changes during the 20th century throughout Portugal, the mean height of males increased from 163.2 cm in 1904 to 172.13 cm with significant differences by district [19]. There is no homogeneity through time or space in Portugal.

5. Conclusions

The main goal of this research is to present a new approach that is applicable in a mass grave context to assess that a specific, relatively homogeneous group was targeted without resorting to harmful typological pseudo-scientific stereotypes. This method may be used in more recent forensic contexts and possibly in some bioarchaeological applications. We used 10 standard cranial measurements that capture overall cranial size and shape and that are relatively resistant to premortem, perimortem, and taphonomic changes. All the measurements can be easily collected with standard spreading and sliding calipers with minimal training, and typically are already collected as part of a standard protocol in many investigations. The results suggest that these standard measurements can be applied to mass graves that have already been excavated, as well as to future investigations. Missing data or measurements for any one case will not compromise the utility of the method, provided that minimum sample sizes are available for analysis. Additional research is required to assess which other metric, morphological, and non-metric variables could reliably be used as predictor variables.

The pattern of results clearly supports that this method does provide useful information for the investigation of mass graves with minimum sample sizes of greater than 25 for a multi-sex scenario and greater than 13 individuals for sex-specific approaches. The results from this research also provide additional evidence against a typological approach in forensic anthropology involving race, ancestry, and affinity. The Coimbra collection subsamples of European-born individuals consistently did not cluster with those individuals from the Terry collection who were described as "White." Those individuals who were described as "White" and "Black" from the Terry collection were consistently grouped together.

References

1. Albanese, J.; Dagdag, A.; Skalic, C.; Osley, S.; Cardoso, H. The fall and rise of identified reference collection: It is possible and necessary to transition from a typological conceptualization of variation to effective utilization of collections. *For. Sci.* **2022**, *2*, 438–454. [CrossRef]
2. Albanese, J.; Saunders, S.R. Is it Possible To Escape Racial Typology In *Forensic Identification*? In *Forensic Anthropology and Medicine: Complementary Sciences from Recovery to Cause of Death*; Schmitt, A., Cunha, E., Pinheiro, J., Eds.; Humana Press: Totowa, NJ, USA, 2006; pp. 281–315.
3. Plens, C.R.; De Souza, C.D.; Albanese, J.; Lopez Capp, T.T.; De Paiva, L.A.S. Reflections on methods to estimate race and ancestry on reference osteological samples in the Brazilian context. *Ethics Med. Public Health* **2021**, *18*, 100680. [CrossRef]
4. Birkby, W.H. An evaluation of race and sex identification from cranial measurements. *Am. J. Phys. Anthropol.* **1966**, *24*, 21–27. [CrossRef] [PubMed]
5. Albanese, J.; Tuck, A.; Gomes, J.; Cardoso, H.F.V. An alternative approach for estimating stature from long bones that is not population- or group-specific. *Forensic Sci. Int.* **2016**, *259*, 59–68. [CrossRef] [PubMed]
6. Digangi, E.A.; Bethard, J.D. Uncloaking a lost cause: Decolonizing ancestry estimation in the United States. *Am. J. Phys. Anthropol.* **2021**, *175*, 422–436. [CrossRef] [PubMed]
7. Albanese, J.; Halliday, J. Muitas maneiras de desaparecer: Defendendo os direitos humanos por meio de padrões éticos e pesquisas mais rigorosas em uma antropologia forense crítica. In *Direito Humanos Sob A Perspectiva Do Direito À Vida, Da Antropologia Forense E Da Justiça No Caso De Violações*; Plens, C., Ed.; Editora Annablume/HHRRC (AAFS): São Paulo, Brazil, 2022; pp. 263–300.
8. Jessee, E.; Skinner, M. A typology of mass grave and mass grave-related sites. *Forensic Sci. Int.* **2005**, *152*, 55–59. [CrossRef] [PubMed]
9. Hunt, D.R.; Albanese, J. The history and demographic composition of the Robert J. Terry Anatomical Collection. *Am. J. Phys. Anthropol.* **2005**, *127*, 406–417. [CrossRef] [PubMed]
10. Albanese, J. Approaches for Constructing and Using Reference Samples to Maximize the Utility of New Forensic Methods. *Eur. J. Anat.* **2021**, *25*, 63–82.
11. Albanese, J. A metric method for sex determination using the hipbone and femur. *J. Forensic Sci.* **2003**, *48*, 263–273. [CrossRef] [PubMed]
12. Rocha, A. Les collections ostéologiques humaines identifiées du Musée Anthropologigue de l'Université de Coimbra. *Antropol. Port.* **1995**, *13*, 7–38.
13. Albanese, J. Strategies for Dealing with Bias in Identified Reference Collections and Implications for Research in the 21st Century. In *Identified Skeletal Collections: The Testing Ground of Anthropology*; Henderson, C.Y., Cardoso, F.A., Eds.; Archaeopress: Oxford, UK, 2018; pp. 59–82.
14. Moore-Jansen, P.H.; Ousley, S.D.; Jantz, R.L. *Data Collection Procedures for Forensic Skeletal Material*; Report No. 48; Department of Anthropology, The University of Tennessee: Knoxville, TN, USA, 1994.
15. Saunders, S.R.; Hoppa, R.D. Sex allocation from long bone measurements using logistic regression. *Can Soc Forens Sci J.* **1997**, *30*, 49–60. [CrossRef]
16. Fausto-Sterling, A. The bare bones of sex: Part 1—Sex and gender. *Signs J. Women in Cult. Soc.* **2005**, *30*, 1491–1527. [CrossRef]
17. Albanese, J. Some Preliminary Finds from the Tholos Tomb and Ossuary at Borzi Hill, Tzannata, Kefalonia: An Example of a Multidisciplinary Approach to Understanding of the Lives and Deaths of Mycenaeans. *Rev. M. Estud. Sobre A Morte Os Mortos Eo Morrer* **2022**, *7*, 74–101.
18. Albanese, J. Albanese, J. A critical Review of The Methodology for The Study of Secular Change Using Skeletal Data. In *Papers in Honour of Michael Spence*; Publication No, 9; Ellis, C., Ferris, N., Timmins, P., White, C., Eds.; Ontario Archaeological Society Occasional: London, ON, Canada, 2010; pp. 139–155.
19. Padez, C. Secular trend in Portugal. *J. Hum. Eco.* **2007**, *22*, 15–22. [CrossRef]