forensic sciences

Article

Evaluating Morphological Methods for Sex Estimation on Isolated Human Skeletal Materials: Comparisons of Accuracies between German and South African Skeletal Collections

Avinash Gupta [1,*], Brendon K. Billings [2], Susanne Hummel [1] and Birgit Grosskopf [1]

[1] Department of Historical Anthropology and Human Ecology, University of Göttingen, 37073 Göttingen, Germany
[2] Faculty of Health Sciences, School of Anatomical Sciences, University of the Witwatersrand, Parktown, Johannesburg 2193, South Africa
* Correspondence: gupta.avinash1992@gmail.com

Abstract: Objectives: The focus of this research is to evaluate the sex estimation methods on isolated human materials by applying morphological methods published in various forensic and anthropological literature on different skeletal series. Materials and Methods: 165 individuals from the 19th to 20th century Inden skeletal series, 252 individuals from the 13th to 14th century Lübeck skeletal series of German ancestry housed at the Department of Historical Anthropology and Human Ecology, the University of Göttingen, Germany, and 161 individuals from the 19th and 20th century of South African African ancestry housed within the Raymond A. Dart collection of modern human skeletons at the University of Witwatersrand, South Africa, with crania, mandibles, and pelves, were assessed. The evaluation criteria are burial information on the Inden series, genetic sex on both the Inden and the Lübeck series, and previous demography on cadavers from the South African African series. Results and Discussion: The sex estimation with cranial traits perform better in Inden and South Africa samples and worse in Lübeck sample. The mandible accuracies for pooled sexes are not exemplary, but the individual traits perform better for males in the Inden, Lübeck, and South Africa samples, except for gonion and angle, which performs better in females. The pelvic traits perform better in the Inden and South Africa samples compared to the Lübeck sample. The statistical tests show that there is a huge difference in the accuracy rates and the performance between both population groups from Germany itself, considering that Inden and Lübeck samples share the same ancestry. The accuracy rates improve with the exclusion of ambiguous individuals.

Keywords: sex estimation; forensic anthropology; accuracy rate; morphological traits

check for updates

Citation: Gupta, A.; Billings, B.K.; Hummel, S.; Grosskopf, B. Evaluating Morphological Methods for Sex Estimation on Isolated Human Skeletal Materials: Comparisons of Accuracies between German and South African Skeletal Collections. *Forensic Sci.* **2022**, *2*, 574–584. https://doi.org/10.3390/forensicsci2030042

Academic Editors: Ricardo Dinis-Oliveira, Francisca Alves Cardoso and Pier Matteo Barone

Received: 29 July 2022
Accepted: 24 August 2022
Published: 30 August 2022

1. Introduction

Biological profile measures such as sex, age, stature, and ancestry are important for forensic identification. Sex estimation plays a significant role in relation to the other measures, i.e., age, stature, ancestry, and pathology. By estimating the sex of the individual, the number of possible matches is halved [1]. In principle, for sex estimation three approaches are available—morphological (qualitative and quantitative), metrical (quantitative), and molecular [2]. Heavy reliance of the scientific community on the molecular approach is viewed to be the key component for reducing errors and producing a conclusive result [3]. Even if the skeleton is well preserved, the material might have too highly-degraded DNA [4] for it to be applied. Moreover, DNA analysis is an expensive resource for some countries, which might not have the infrastructure or skills. An Interpol review shows that DNA is used in police investigations in only 84 countries [5]. In such cases, they have to utilize the classical methods, which are morphological and metrical. However, the morphological methods or sexually dimorphic nonmetric traits appear to be more accurate in one population group and not in the others. Although these nonmetric traits are accurate, with 70% to 100% accuracy rates, based on literature published over more than a century [6–8], it could be argued that some methods achieve a high level (100%) of accuracy due to the fact that the populations being tested are similar to those collections used to create the methods or are even the same anatomical collections [9], such as the Robert J. Terry or Raymond A. Dart collections, which are often studied and cited in the literature.

On one hand, the community of forensic anthropology is trying to devise novel metrical methods that are population specific and derived from newer collections [10–15], or from cadaveric collections that are constantly being updated, such as the Dart collection. On the other hand, for morphological methods the sex estimation standards are heavily based on either the 20th century Terry, Herman–Todd, and William Bass collections or war-dead samples such as those from the Korean War and Balkan conflict, with a bias towards one sex [9]. Finding collections with documented accounts of biological sex is difficult, and to evaluate the traits to see if they can be transferred onto other known-sex skeletal series is even more complicated. Sexual dimorphism and the methods of morphological assessment on the skull and pelvis are often based on the differences caused by shape and size; they can either be differences between biological sexes based on the fact that females have different pelvic morphology to accommodate parturition, or other differences of robustness that are prominent in males, for example, the skull [16]. The fact that population variation plays an important role in skeletal sexual dimorphism [17], the groups vary in terms of their activity patterns [2], and there is an implication of secular trends on sex estimation [18], this demonstrates that the use of sex estimation accuracies are affected by the superimposition of the use of one sample onto another sample derived from different collections representing two separate population groups or from the same collection with specimens from a different temporal period.

The idea of the use of population-specific formulae concerning robusticity, degree of sexual dimorphism, and body size [19] leaves plenty of room to check the reliability of universal sets of traits by assessing the populations that were not considered in their study. It may act as a strong argument in the debate on whether it is necessary to develop population-specific data or not. Because the validation studies on sex estimation are based on well-documented and sometimes demographically biased collections [8], it is not clear how reliable the morphological traits are if they are applied to medieval samples. Ancient DNA (aDNA) testing is a widely known tool for profiling archaeological material and estimating sex [20]. Short Tandem Repeats (STR) typing of samples includes PCR approaches that target amelogenin genes with X–Y homology [21]. The success of STR typing depends on the preservation of aDNA. It is important to point out that the accuracy of aDNA methods is high, but there are limitations. These include Y-allele deletion [22,23], leading to false negative results for males [24]; contamination in sample preparation and extraction is difficult to control [25], and most importantly it is an expensive method for a routine analysis by archaeologists or forensic anthropologists. The heavy reliance on DNA and the idea that it will eventually resolve the case is the reason unrecognizable remains are often not fully assessed with traditional anthropological methods, whether metrical or morphological [3]. These limitations justify the need to evaluate the empirical accuracy rates produced by morphological estimation techniques as published in the forensic anthropological literature.

The aim of this article is to evaluate morphological traits for their accuracy and reliability when applied on samples from European and African population groups of the modern and medieval periods. To achieve this, morphological traits will be used to estimate sex for Inden (modern), Lübeck (medieval), and South African (modern) population groups. Altogether, this large-scale analysis with three different population groups will test if standard morphological methods produce a reliable accuracy on modern and medieval population groups from Germany and only modern in South Africa. By evaluating the current morphological approach, this study can highlight which traits should be given more weight and which should be excluded, while establishing universal traits that can be applied irrespective of the population specificity.

2. Materials: From Inden, Lübeck, South Africa

2.1. Inden Collection

Between 1999 and 2004, excavations carried out under the supervision of Dr. Birgit Grosskopf [26] at a cemetery of an old church, St. Clemens, in an Alt-Inden village in Düren district, recovered 236 individuals. The skeletons of the Inden collection come from the period 1877 to 1924 and is housed at the Department of Historical Anthropology and Human Ecology, the University of Göttingen, Germany. The sample for this research includes 165 individuals with compromised preservation of some bones due to taphonomic processes. Out of 109 previously known-sex individuals, 32 individuals had sex determination carried out by genetic analysis.

2.2. Lübeck Collection

The mass graves excavated between 1989 and 1992 in the northern German city of Lübeck include the data of 1255 individuals. Out of 1044 complete and partially complete individuals housed at the Department of Historical Anthropology and Human Ecology, only 252 individuals have been considered. The collection belongs to the Black Death period of 1260 to 1390 AD [27]. The burial information register, or church book, reveals the sex of individuals up to a certain limit. STR typing has been performed for a selection of samples from the Inden and Lübeck series. Together, the church book and genetic analysis provide sex for only 109 individuals for evaluation and accuracy calculation in the Inden series. The Lübeck collection is not demographically known, and genetic estimation performed at the University of Göttingen for 76 individuals are used for accuracy calculation.

2.3. South African Collection

Lastly, the third sample involves individuals from the Raymond A. Dart collection, with four major self-identified groups according to the census data, which include South African (SA) African, SA White, SA Colored, and SA Indian/Asian. SA African (SAA) or Black South African (as per South Africa's official classification for population groups) has multiple tribal affinities, namely nine [28]. The collection housed at the School of Anatomical Sciences, the University of the Witwatersrand, South Africa, is demographically known and belongs to the period of birth years 1827–1980. The total sample size from the Dart collection for this research consists of 161 cadaver-origin skeletons of SA African ancestry with mixed tribal affinities [29]. Sex estimation is carried out based on the evaluation of individual morphological traits, as well as a combination of traits. Because the accuracy of aDNA is much higher than for morphological traits, this baseline will be used to compare results and calculate the accuracy rates, at least in the Inden and Lübeck series, because the demography of the South African series is already known.

2.4. Criteria for Selection of Bones

The focus of this study was on only the following bones: crania, mandible, and pelvis. The pelvis is generally considered to be the most sexually dimorphic due to parturition [19,30], which makes it an integral skeletal region to be considered in any morphological sex estimation study. When the pelvis is fragmented or absent, the skull follows the standard framework of analysis among other bones [30]. Long bones offer a better prediction based on dimensions [31], which can be obtained using metrical methods, hence they are excluded from this study.

3. Methods

3.1. Morphological Sex Estimation Methodology

This was a blind study, which means that the sex estimation of the individuals was first completely performed in a morphological series, without any prior knowledge regarding the sex of the individuals. It requires a significant amount of experience to correctly evaluate skeletal material with morphological traits. Thus, the experiment was conducted under the supervision of a more experienced forensic anthropologist. The four chosen traits of the innominate, as shown in Table 1, reflect the morphology of pubic, ischial, and iliac bones. The nine cranial and five mandibular traits mentioned in Tables 2 and 3 have been observed, evaluated, and estimated for a particular sex, as proposed by the authors [32–38]. For crania, mandible, and pelvis, the traits were observed independent from each other, either as a single trait or as a group of two traits together.

Table 1. Morphological pelvic traits used for sex estimation.

Traits	Common Sources
Greater sciatic notch Arc composé (Composite arch) Sub-pubic angle Iliac crest	(Bruzek, 2002; Buikstra and Ubelaker, 1994; Ferembach, 1980; Grupe et al., 2015; Klepinger, 2006; Novotný, 1981; Hermann et al., 1990) [32–38]

Table 2. Morphological mandibular traits used for sex estimation.

Traits	Common Sources
Mandible overall Condylar process Mentum (mental eminence) Gonion and angle Corpus height	(Buikstra and Ubelaker, 1994; Ferembach, 1980; Grupe et al., 2015; Herrmann et al., 1990; Klepinger, 2006; Loth & Henneberg, 1996) [33–37,39]

Table 3. Morphological cranial traits used for sex estimation.

Traits	Common Sources
Cranium overall Frontal tuberosity (eminence) and forehead steepness Glabella and supraorbital ridges (superciliary arch) Eye orbitals Zygomaticum (zygomatic arch) Margo orbitalis (supraorbital margin) Mastoid process	(Buikstra and Ubelaker, 1994; Ferembach, 1980; Grupe et al., 2015; Herrmann et al., 1990; Klepinger, 2006) [33–37]

3.2. Morphological Series

The comparison of the shapes of traits forms the basis of a morphological series, where pronounced or robust male and female characteristics form the reference. This reference was used to observe and evaluate the remaining individuals. Morphological series is advantageous for individuals with underdeveloped or less-pronounced traits. They can be placed in a range, easily comparing the features with other individuals. The five-point scale from the Workshop of European Anthropologists [34] includes scoring of −2 (hyperfeminine) to 0 (neutral) to +2 (hypermasculine). This scoring was modified for this research. The observation was based on a particular range, which is Male, Tendency Male, Indifferent, Tendency Female, Female. When there was a strong fragmentation in an individual for a particular trait, it was recorded as indeterminate. Indeterminate is not a part of the five-point range. Classifying an individual as indeterminate would exclude the individual, but the same individual was included in the study for other observable traits or intact bony landmarks. When features extended towards both male and female, the sex was estimated as indifferent.

3.3. Known Sex and Morphological Sex Estimation Comparison: Calculation of Accuracy Rate

Accuracy rate is the percentage of correctly classified individuals in the corresponding groups of male and female. To calculate this, the sex estimated by the morphological assessment is compared with the documented biological sex. In order to perform the comparison, the morphological estimate Tendency Male was categorized under Male and Tendency Female under Female. For Inden, Lübeck, and the South African collections, the following information was available on their demography—Church book records and molecular sex estimation for Inden; cadaver-origin skeletons from SA had a database maintained locally where most of the individuals had correct sex defined from the death certificate. In some instances, the sex was incorrect, which could be a limitation to the study. The incorrectly-sexed individuals are few. The ancestry estimation for those individuals might be difficult, but sex is more easily identified from a cadaver that is from an unclaimed provenance, which is why those individuals have still been included in the study based on their identified sex. The Lübeck series had data from the molecular sex estimation performed at the department of Historical Anthropology, the University of Göttingen. To find the accuracy rate, the number of correctly classified individuals is divided by the total number of individuals, which were morphologically assessed.

$$\text{Rate of accuracy } (\%) = \frac{The\ number\ of\ correctly\ classified\ individuals}{Total\ number\ of\ individuals} \times 100 \quad (1)$$

For the calculation of accuracy rate, 'indeterminate' (broken or unobservable or not available) individuals have not been included in the study but 'indifferent' (or ambiguous) ones are already included. Inskip et al. (2019) and McFadden and Oxenham (2016) [9,40] calculated accuracy rates by including ambiguous sex estimates. This raises an interesting argument that indifferent individuals should be left out of accuracy calculations because ambiguity cannot be taken as incorrect estimation; the counter argument justifies the inclusion of indifferent individuals because they help quantify the information on the high number of indifferent sex estimates. This sheds light on the trait sexual dimorphism power [9]. In the results section, two accuracy rates are mentioned, one termed as raw accuracy with 'indifferent' individuals and the second without them. This aids in refraining from reporting the overestimated accuracy, contrary to cases where a researcher might include only correctly-sexed individuals, showing that a particular method for sex estimation has a very high accuracy rate.

The significance of the results was assessed by two tests, namely, the McNemar test and Cohen's Kappa test. The McNemar test assesses the systematic differences between the conclusion made by previously known sex and the morphological sex estimation, whereas Cohen's Kappa shows the agreement between the known sex and morphological estimation. The level of disagreement would be higher if the McNemar values would be lower than 1. The level of agreement for Cohen's Kappa is assessed, as shown in the following range, taken from Watson and Petrie (2010)—"Poor if $k < 0.00$, slight if $0.00 \leq k \leq 0.20$, fair if $0.21 \leq k \leq 0.40$, moderate if $0.41 \leq k \leq 0.60$, substantial if $0.61 \leq k \leq 0.80$, and almost perfect if $k > 0.80$" [41].

4. Results

The number of individuals in all three skeleton collections identified by morphological methods and previously known sex in the form of genetic analyses, burial information, and cadaver database, is outlined in Table 4. The numbers in the table do not exclude the skeletons with unobservable features for individual traits. This gives an overview of the sample size for three population groups.

Table 4. Number of individuals by morphologically estimated and previously known sex.

	Population Groups		
Source of Sex Estimate	**Inden**	**Lübeck**	**South Africa**
Morphological	164	236	161
Previously known	109	76	161

Note. The previously known data include the sources aDNA analysis, church book, and the cadaver database.

The accuracy rates (with and without ambiguous individuals) and Cohen's Kappa results for individual traits for the Inden series is shown in Table 5, for the Lübeck series in Table 6, and for the South African series in Table 7. For crania in the Inden and SA groups, glabella-supra orbital ridges present the highest accuracy. Although in the Lübeck series the combination of traits as overall assessment performs better than other traits, the small sample size impacts greatly in terms of poor accuracy rates. Additionally, Cohen's Kappa tests demonstrate a similar pattern that there is a higher level of agreement for traits in Inden and SA but not for the Lübeck series. None of the series have any traits that show a perfect agreement (k = 1.000). Glabella-supra orbital ridges and cranium overall show a substantial agreement in Inden and South Africa with previously known sex, whereas the Lübeck series has the poorest agreement between morphological sex estimates and previously known sex estimates for all the cranial traits. McNemar tests identify the systematic difference between morphological and previously known sex estimates, which shows a trend in the Lübeck cranial traits.

Table 5. Total number of individuals for skull and os coxae with traits observed: the correctly classified, indifferent (ambiguous), and incorrect results and the corresponding accuracy rates in the Inden series.

Trait	Individuals Considered with Known Sex	Match	No Match	Raw Accuracy (%)	Indifferent	Without Indifferent	Match	No Match	Accuracy (%)	Cohen's Kappa	*p*-Value
Inden Calvarium—Morphological											
Cranium overall	74	51	23	68.92	13	61	51	10	83.61	0.665	0.000
Frontal tuberosity and Steepness	75	59	16	**78.7**	13	62	52	10	83.9	0.665	0.000
Glabella-supra orbital ridges	75	53	22	70.7	12	63	53	10	84.1	0.679	0.000
Margo orbitalis	75	43	32	57.3	17	58	43	15	74.1	0.450	0.000
Eye orbitals	62	43	19	69.4	4	58	43	15	74.1	0.464	0.000
Zygomaticum	69	44	25	63.8	9	60	44	16	73.3	0.444	0.000
Mastoid process	75	43	32	57.3	13	62	42	20	67.7	0.351	0.000
Inden Mandible—Morphological											
Mandible overall	65	50	15	**76.9**	0	65	50	15	76.9	0.491	0.000
Condylar process	64	43	21	67.2	0	64	43	21	67.2	0.294	0.019
Corpus height	59	39	20	66.1	0	59	39	20	66.1	0.264	0.040
Mentum	57	40	17	70.2	0	57	40	17	70.2	0.348	0.006
Gonion and angle	54	35	19	64.8	1	53	35	18	66.0	0.183	0.052
Inden Pelvis—Morphological											
Iliac crest	69	47	22	68.1	0	69	47	22	68.1	0.361	0.003
Arc compose	90	67	23	74.4	3	87	67	20	77.0	0.530	0.000
Greater schiatic notch	90	65	25	72.2	7	83	65	18	78.3	0.556	0.000
Sub pubic angle	54	48	6	**88.9**	1	53	48	5	90.6	0.797	0.000

Note: The numbers in bold highlight the highest accuracy among all the traits for a particular skeletal element.

In the Inden and South African groups, mandible performs the best overall, considering they have different sample sizes. For pelvic traits, sub-pubic angle is the best indicator of sex in Inden and SA, and is comparably better in Lübeck. The best traits for crania have lower accuracy than the accuracy achieved by the best pelvic trait in all three series. Highlighting the indifferent or ambiguous individuals, the highest variation could be seen in cranial traits by looking at the difference between the raw and the general accuracy rates. The McNemar test shows that the most systematic difference between all traits in Inden is gonion and angle (McN = 0.000); this highlights the difference between correct and incorrect classifications. This systematic difference is quite high for the cranial traits (0.000 to 0.002) in Lübeck. Cohen's Kappa results suggest that the pelvic trait sub-pubic angle has an almost perfect agreement between morphological and previously known-sex estimates in Inden (k = 0.797) and SA (k = 0.879).

Table 6. Total number of skull and os coxae traits observed: the correctly classified, indifferent (ambiguous), and incorrect results and the corresponding accuracy rates in the South African series.

Trait	Individuals Considered with Known Sex	Match	No Match	Raw Accuracy (%)	Indifferent	Without Indifferent	Match	No Match	Accuracy (%)	Cohen's Kappa	*p*-Value
South Africa Calvarium—Morphological											
Cranium overall	160	130	30	81.3	0	160	130	30	81.3	0.595	0.000
Frontal tuberosity and Steepness	160	125	35	78.1	2	158	125	33	79.1	0.573	0.000
Glabella-supra orbital ridges	160	130	30	81.3	1	159	130	29	81.8	0.613	0.000
Margo orbitalis	160	108	52	67.5	1	159	108	51	67.9	0.299	0.000
Eye orbitals	160	124	36	77.5	2	158	124	34	78.5	0.533	0.000
Zygomaticum	160	124	36	77.5	4	156	124	32	79.5	0.581	0.000
Mastoid process	160	113	47	70.6	1	159	113	46	71.1	0.402	0.000
South Africa Mandible—Morphological											
Mandible overall	160	130	30	**81.3**	4	156	130	26	83.3	0.634	0.000
Condylar process	160	114	46	71.3	1	159	114	45	71.7	0.407	0.000
Corpus height	160	117	43	73.1	4	156	117	39	75.0	0.464	0.000
Gonion and angle	160	127	33	79.4	2	158	127	31	80.4	0.604	0.000
Mentum	160	102	58	63.8	5	155	102	53	65.8	0.290	0.000
South Africa Pelvis—Morphological											
Iliac crest	156	109	47	69.9	1	155	109	46	70.3	0.382	0.000
Greater schiatic notch	156	122	34	78.2	2	154	122	32	79.2	0.552	0.000
Arc compose	156	127	29	81.4	1	155	127	28	81.9	0.609	0.000
Sub pubic angle	156	146	10	93.6	1	155	146	9	94.2	0.879	0.000

Note: The numbers in bold highlight the highest accuracy among all the traits for a particular skeletal element.

Table 7. Total number of skull and os coxae traits observed: the correctly classified, indifferent (ambiguous), and incorrect results and the corresponding accuracy rates in the Lübeck series.

Trait	Individuals Considered with Known	Match	No Match	Raw Accuracy (%)	Indifferent	Without Indifferent	Match	No Match	Accuracy (%)	Cohen' s Kappa	*p*-Value
Lübeck Calvarium—Morphological											
Cranium overall	43	23	20	**53.5**	4	39	23	16	59.0	0.273	0.027
Frontal tubercle and Steepness	40	21	19	52.5	3	37	21	16	56.8	0.204	0.121
Glabella-supra orbital ridges	43	20	23	46.5	8	35	20	15	57.1	0.274	0.018
Margo orbitalis SH	42	18	24	42.9	7	35	18	17	51.4	0.177	0.128
Eye orbitals	43	23	20	53.5	1	42	23	19	54.8	0.222	0.045
Zygomaticum	43	20	23	46.5	6	37	20	17	54.1	0.209	0.076
Mastoid process	43	20	23	46.5	6	37	20	17	54.1	0.211	0.074
Lübeck Mandible—Morphological											
Mandible overall	31	18	13	**58.1**	0	31	18	13	58.1	0.074	0.675
Condylar process	26	12	14	46.2	1	25	12	13	48.0	0.110	0.405
Corpus height	31	17	14	54.8	0	31	17	14	54.8	0.084	0.609
Mentum	31	13	18	41.9	1	30	13	17	43.3	−0.049	0.745
Gonion and angle	31	22	9	71.0	0	31	22	9	71.0	0.318	0.076
Lübeck Pelvis—Morphological											
Iliac crest	51	31	20	60.8	7	44	31	13	70.5	0.441	0.001
Arc compose	51	40	11	**78.4**	3	48	40	8	83.3	0.652	0.000
Greater schiatic notch	50	36	14	72.0	7	43	36	7	83.7	0.670	0.000
Sub pubic angle	51	32	19	62.7	13	38	32	6	84.2	0.687	0.000

Note: The numbers in bold highlight the highest accuracy among all the traits for a particular skeletal element.

Table 8 highlights the distribution of accuracy rates amongst males and females. It shows a comparison of trait classification in all the population groups. For each trait in every population group, there are values in Male that match and in Female that match. The value inside the brackets is the total number of individuals for a particular sex, and the value outside the brackets shows the correctly identified individuals. The pubic traits were the best indicators for male and female identification in all three series. The probability of males being scored or categorized as male, and females being scored or categorized as female is high for subpubic angle. For South African females, it even reaches 95%, which is far above the pooled accuracy rates in every series. For skull traits in the Inden series, the males were identified poorly with the mastoid process, with the lowest accuracy rate, and in the Inden females the zygomaticum or zygomatic arch was the worst sex indicator. In the Lübeck series, although the sample size is low, a very stark situation can be observed for glabella-supra orbital ridges in females, securing only 28.6% as opposed to the male–female pooled accuracy rate in the Inden series (84.3%), and in the Inden female it can be seen as the best performing sex indicator with 85.2% accuracy.

Table 8. Accuracy rates for individual traits in males and females separately.

Trait	Inden					Lübeck					South Africa				
	Male Match	**Male Accuracy (%)**	**Female Match**	**Female Accuracy (%)**	**McNemar Test**	**Male Match**	**Male Accuracy (%)**	**Female Match**	**Female Accuracy (%)**	**McNemar Test**	**Male Match**	**Male Accuracy (%)**	**Female Match**	**Female Accuracy (%)**	**McNemar Test**
Cranium—Morphological															
Cranium overall	30 (36)	84	21/25	83.30	0.754	11 (12)	91.7	12 (27)	44.40	0.001	87 (98)	**88.8**	43 (62)	69.4	0.200
Frontal tubercle and Steepness	32 (37)	**86.5**	20 (25)	80	1.000	8 (10)	80	13 (27)	**48.1**	0.004	75 (96)	78.1	50 (62)	**80.60**	0.163
Glabella-supra orbital ridges	30 (36)	83.3	23 (27)	**85.2**	0.754	11 (11)	**100**	9 (24)	37.5	0.000	84 (97)	86.6	46 (62)	74.2	0.711
Margo orbitalis	29 (34)	85.3	14 (24)	58.3	0.302	10 (11)	90.9	8 (24)	33.3	0.000	78 (97)	80.4	30 (62)	48.4	0.092
Eye orbitals	16 (24)	66.7	27 (34)	79.4	1.000	11 (12)	91.7	12 (30)	40	0.000	85 (96)	88.5	39 (62)	62.9	0.058
Zygomaticum	28 (36)	77.8	16 (24)	66.7	1.000	11 (12)	91.7	9 (25)	36	0.000	74 (94)	78.7	50 (62)	80.6	0.215
Mastoid process	24 (36)	66.7	18 (26)	69.2	0.503	10 (11)	90.9	10 (26)	38.5	0.000	71 (97)	73.2	42 (62)	67.7	0.461
Mandible—Morphological															
Mandible overall	35 (41)	85.4	15 (24)	**62.5**	0.607	4 (9)	44	14 (22)	63.6	0.581	89 (96)	**92.7**	41 (60)	68.3	0.029
Condylar process	30 (40)	75	13 (24)	54.2	1.000	4 (5)	**80**	8 (20)	40	0.003	74 (97)	76.3	40 (62)	64.5	1.000
Corpus height	28 (36)	77.8	11 (23)	47.8	0.503	5 (9)	55.6	12 (22)	54.5	0.180	79 (95)	83.2	38 (61)	62.3	0.337
Mentum	29 (34)	85.3	11 (23)	47.8	0.143	5 (9)	55.6	8 (21)	38.1	0.049	66 (95)	69.5	36 (60)	60	0.011
Gonion and angle	31 (32)	**96.9**	4 (21)	19	0.000	5 (9)	55.6	17 (22)	**77.3**	1.000	74 (97)	90.2	53 (61)	**86.9**	0.583
Pelvis—Morphological															
Iliac crest	26 (39)	66.7	21 (30)	70	0.523	14 (15)	93.3	17 (29)	58.6	0.003	70 (95)	73.7	39 (60)	65	0.659
Arc compose	40 (49)	81.6	27 (38)	71.1	0.824	15 (18)	83.3	25 (30)	**83.3**	0.727	85 (96)	88.5	42 (59)	71.2	0.215
Greater schiatic notch	39 (47)	83	26 (36)	72.2	0.815	15 (17)	88.2	21 (26)	80.8	0.453	82 (94)	87.2	40 (60)	66.7	0.345
Sub pubic angle	31 (33)	**93.9**	17 (20)	**85**	1.000	15 (16)	**93.8**	17 (22)	77.3	0.219	89 (95)	**93.7**	57 (60)	**95**	0.508

Note: The numbers in bold highlight the highest accuracy among all the traits for a particular skeletal element.

5. Discussion

The aim of this study was to evaluate the morphological traits for their accuracy and reliability on modern and medieval samples from European ancestry and on modern sample from African ancestry. The sex estimation for an individual can be performed either based on individual trait assessment or on multiple trait assessment, methods, and skeletal regions combined. Looking at the accuracy rates of the assessed traits in Table 8 on all three population groups, it can be seen that the sex can be accurately estimated if all the pelvic traits are observable and can perform even better if correctly sexed in conjunction with skull traits. This finding is congruent with the study performed by Inskip et al. (2019) [9]; the accuracy rates of the morphological traits in an English skeletal collection are higher for pooled sexes (95.6%) if skull and os coxae are assessed together (with 95.6% raw accuracy) and it outperforms the assessment of os coxae alone, which achieved only 91.8% accuracy. When a forensic anthropologist performs analysis on skeletal elements, "ideally the pelvis should be assessed first, because its evolutionary predisposition to parturition and absence of ancestral traits lead to greater reliability than is possible from assessing the skull, which should be assessed second" [42,43]. This statement appears to be more valid for subpubic angle than the other pelvic traits, as seen for the Inden and South Africa series. The difference between males and females' accuracies is quite moderate for arc compose, greater sciatic notch, and iliac crest in all three groups. The higher accuracy rates in Inden and South African males is comparable to the study carried out by Đurić et al. [6]. They found over 90% accuracy for arc compose when they assessed only males of the Balkan group.

If pelvic bones are fragmented or broken, the misclassifications based only on the skull would be even greater. As shown in Table 5, there is a slight difference in male and female accuracies for sub pubic angle in the Inden series (male = 93.9%, female = 85%), but if skull traits are assessed independent of pelvic traits, the number of misclassified individuals appears to be high among Lübeck and South African females for most of the skull traits, which can be clearly seen in Table 8. In the pelvis, the variation between males and females is related to reproduction, and in crania the sexual dimorphism is more related to body size and shape, which highlights musculature and robusticity [44]. This means it has more potential to vary among different population groups [9]. Inden and Lübeck are both German skeletal collections, but from different periods of modern and medieval. The stark difference between the accuracy rates for cranial traits in both series, as seen in Tables 5 and 7, is supported by Inskip et al. (2019) [9]. The authors sourced traits from papers that used modern collections, to which they compared the results obtained from medieval English samples. It showed a higher degree of differences for sexual dimorphism. This research shows that independent traits might not give an appropriate sex estimation when the standards from modern collections are applied on archaeological samples.

The skull sex estimates, and previously determined sex, have a substantial agreement and an improved accuracy rate when ambiguous individuals are excluded. This stark change can be observed in the Inden series, particularly for cranium overall, glabella-supraorbital ridges, and margo orbitalis. It was noted that the skull is highly variable from the data, but it also scored high accuracies for overall traits than for individual traits. The difference between raw and correctly-sexed accuracies highlights that there are more ambiguous individuals for skull traits than for pelvic traits and more overlapping in Inden than in Lübeck and South Africa. This is due to the fact that both males and females of the Inden skeletal collection present with more robust and prominent features in general, and thus there is more overlap. An additional reason could be that development of features in males and females take place at different ages. The masculine features develop at an extended period of growth; male individuals who die at a young age might appear feminine. Females might show more robust features as they age [33]. This could become a factor for ambiguity and misclassifications in Inden, Lübeck, and SA females. The females might be of older age groups and hence are often classified as males. In light of this, it is important to include a larger sample size with more females of different age groups.

6. Conclusions

The modern collections of European and African ancestries score better accuracies when using morphological methods based on modern collections when compared to archaeological collections. The cranial traits, combined and assessed overall, perform as one of the best in all three series, namely, Inden, Lübeck, and SA groups, despite being of different ancestries and temporal periods. The small sample size in every series has been a big limitation of this study, and this may create a bias. For every trait tested, the sex difference on accuracy rate could stem from the unequal numbers of male and females in each skeletal series. This greatly impacts the accuracy rates of all the traits in the Lübeck series, which do not improve much, even after taking the ambiguous individuals out. If researchers are excluding ambiguous or indifferent features to ascertain higher accuracies in sex estimation, this provides a significant research theme for the future. The combination of traits for mandible region follows the same pattern, achieving the highest accuracy in all three series. The strongest individual trait in the pelvis turns out to be the subpubic angle, which scores the highest accuracy in males and females of modern series Inden and SA, and also in the males of archaeological samples of Lübeck, and this trait could be considered as a universal trait that can be applied, regardless of the population specificity. Interestingly enough, the females in Lübeck perform better with arc compose. The performances of different individual traits proves that a partially complete or fragmented pelvis would provide a clearer picture on the biological sex of an individual than the fragmented skull, which would be more useful if all the traits are assessed together. The Lübeck series does not deliver the same accuracy rate as the other two modern samples from the Inden and SA series when using the cranial and mandibular features, which may be related to the skull variation in different population groups. Krüger et al. (2015) address the accuracy of morphological traits on Black and White South African crania. According to them, the expression of morphological cranial traits vary across populations where the classification rate or accuracy is highly dependent on the reference collection that is used to compare unidentified skulls [45]. The inability to easily identify the morphology on Lübeck archaeological sample emphasizes that modern samples have better preserved bones and associated morphological features.

Author Contributions: Conceptualization, methodology, original draft preparation, A.G., B.G., S.H. and B.K.B.; formal Analysis, investigation, resources, data curation, project administration, A.G.; review and editing, B.G., S.H. and B.K.B.; visualisation and supervision, B.G. and S.H.; funding acquisition, A.G., B.G. and S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Academic Exchange Service (Deutscher Akademischer Austauschdienst) research scholarship for doctoral students with funding program number 57299294.

Institutional Review Board Statement: Ethical review and approval were waived for this study as it includes human skeletal material that have been retrieved from burials and mass graves in Germany. The request of the ethics waiver for accessing the Raymond Dart cadaver collection has been approved by the School of Anatomical Sciences, University of Witwatersrand.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the study's conclusions are available on reasonable request from the corresponding author.

Acknowledgments: The authors appreciate the reviewers' insightful comments, which significantly enhanced the paper. Special thanks to the Department of Historical Anthropology in Germany and the Raymond Dart collection in South Africa for providing the skeletal material for this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Mokoena, P.; Billings, B.K.; Gibbon, V.; Bidmos, M.A.; Mazengenya, P. Development of Discriminant Functions to Estimate Sex in Upper Limb Bones for Mixed Ancestry South Africans. *Sci. Justice* **2019**, *59*, 660–666. [CrossRef] [PubMed]
2. Krishan, K.; Kanchan, T.; Kharoshah, M.A. "Advances in Forensic Anthropology"–Creation of Skeletal Databases for Forensic Anthropology Research and Casework. *Egypt. J. Forensic Sci.* **2016**, *6*, 29–30. [CrossRef]
3. de Boer, H.H.; Obertová, Z.; Cunha, E.; Adalian, P.; Baccino, E.; Fracasso, T.; Kranioti, E.; Lefévre, P.; Lynnerup, N.; Petaros, A. Strengthening the Role of Forensic Anthropology in Personal Identification: Position Statement by the Board of the Forensic Anthropology Society of Europe (FASE). *Forensic Sci. Int.* **2020**, *315*, 110456. [CrossRef] [PubMed]
4. Hummel, S. *Ancient DNA Typing*; Springer: Berlin, Germany, 2003.
5. Butler, J.M.; Willis, S. Interpol Review of Forensic Biology and Forensic DNA Typing 2016-2019. *Forensic Sci. Int. Synerg.* **2020**, *2*, 352–367. [CrossRef]
6. Đurić, M.; Rakočević, Z.; Đonić, D. The Reliability of Sex Determination of Skeletons from Forensic Context in the Balkans. *Forensic Sci. Int.* **2005**, *147*, 159–164. [CrossRef]
7. Lewis, C.J.; Garvin, H.M. Reliability of the Walker Cranial Nonmetric Method and Implications for Sex Estimation. *J. Forensic Sci.* **2016**, *61*, 743–751. [CrossRef]
8. Thomas, R.M.; Parks, C.L.; Richard, A.H. Accuracy Rates of Sex Estimation by Forensic Anthropologists through Comparison with DNA Typing Results in Forensic Casework. *J. Forensic Sci.* **2016**, *61*, 1307–1310. [CrossRef]
9. Inskip, S.; Scheib, C.L.; Wohns, A.W.; Ge, X.; Kivisild, T.; Robb, J. Evaluating Macroscopic Sex Estimation Methods Using Genetically Sexed Archaeological Material: The Medieval Skeletal Collection from St John's Divinity School, Cambridge. *Am. J. Phys. Anthropol.* **2019**, *168*, 340–351. [CrossRef]
10. Gualdi-Russo, E. Sex Determination from the Talus and Calcaneus Measurements. *Forensic Sci. Int.* **2007**, *171*, 151–156. [CrossRef]
11. Guyomarc'h, P.; Bruzek, J. Accuracy and Reliability in Sex Determination from Skulls: A Comparison of Fordisc® 3.0 and the Discriminant Function Analysis. *Forensic Sci. Int.* **2011**, *208*, 180.e1–180.e6. [CrossRef]
12. Trautmann, M.; Trautmann, I.; Hotz, G. A Simple Metric Sexing Method for Unknown Skeletal Remains: The Sacro-Clavicular Index (SCI). *Anthropol. Anz.* **2014**, *71*, 57–64. [CrossRef]
13. Selliah, P.; Martino, F.; Cummaudo, M.; Indra, L.; Biehler-Gomez, L.; Campobasso, C.P.; Cattaneo, C. Sex Estimation of Skeletons in Middle and Late Adulthood: Reliability of Pelvic Morphological Traits and Long Bone Metrics on an Italian Skeletal Collection. *Int. J. Leg. Med.* **2020**, *134*, 1683–1690. [CrossRef]
14. Introna, F.; Di Vella, G.; Campobasso, C.P.; Dragone, M. Sex Determination by Discriminant Analysis of Calcanei Measurements. *J. Forensic Sci.* **1997**, *42*, 725–728.
15. Introna, F.; Di Vella, G.; Campobasso, C.P. Sex Determination by Discriminant Analysis of Patella Measurements. *Forensic Sci. Int.* **1998**, *95*, 39–45. [CrossRef]

16. Glucksmann, A. *Sexual Dimorphism in Human and Mammalian Biology and Pathology*; Academic Press: Cambridge, MA, USA, 1981.
17. Ubelaker, D.H.; DeGaglia, C.M. Population Variation in Skeletal Sexual Dimorphism. *Forensic Sci. Int.* **2017**, *278*, 407.e1–407.e7.
18. Ross, A.H.; Ubelaker, D.H.; Kimmerle, E.H. Implications of Dimorphism, Population Variation, and Secular Change in Estimating Population Affinity in the Iberian Peninsula. *Forensic Sci. Int.* **2011**, *206*, 214.e1–214.e5. [CrossRef]
19. Steyn, M.; Patriquin, M.L. Osteometric Sex Determination from the Pelvis—Does Population Specificity Matter? *Forensic Sci. Int.* **2009**, *191*, 113.e1–113.e5. [CrossRef]
20. Seidenberg, V.; Schilz, F.; Pfister, D.; Georges, L.; Fehren-Schmitz, L.; Hummel, S. A New MiniSTR Heptaplex System for Genetic Fingerprinting of Ancient DNA from Archaeological Human Bone. *J. Archaeol. Sci.* **2012**, *39*, 3224–3229.
21. Haas-Rochholz, H.; Weiler, G. Additional Primer Sets for an Amelogenin Gene PCR-Based DNA-Sex Test. *Int. J. Leg. Med.* **1997**, *110*, 312–315.
22. Schmidt, D.; Hummel, S.; Herrmann, B. Brief Communication: Multiplex X/Y-PCR Improves Sex Identification in ADNA Analysis. *Am. J. Phys. Anthropol.* **2003**, *121*, 337–341. [CrossRef]
23. Tozzo, P.; Giuliodori, A.; Corato, S.; Ponzano, E.; Rodriguez, D.; Caenazzo, L. Deletion of Amelogenin Y-Locus in Forensics: Literature Revision and Description of a Novel Method for Sex Confirmation. *J. Forensic Leg. Med.* **2013**, *20*, 387–391. [PubMed]
24. Kashyap, V.K.; Sahoo, S.; Sitalaximi, T.; Trivedi, R. Deletions in the Y-Derived Amelogenin Gene Fragment in the Indian Population. *BMC Med. Genet.* **2006**, *7*, 37.
25. Schmidt, N.; Schücker, K.; Krause, I.; Dörk, T.; Klintschar, M.; Hummel, S. Genome-Wide SNP Typing of Ancient DNA: Determination of Hair and Eye Color of Bronze Age Humans from Their Skeletal Remains. *Am. J. Phys. Anthropol.* **2020**, *172*, 99–109. [CrossRef]
26. Salega, S.; Grosskopf, B. Evaluation of Entheseal Changes in a Modern Identified Skeletal Collection from Inden (Germany). *Int. J. Osteoarchaeol.* **2022**, *32*, 86–99. [CrossRef]
27. Feicke, M. *Inventory and Morphological Sexing in a Medieval Skeletal Series from Lübeck, Department of Historical Anthropology and Human Ecology, Johann-Friedrich-Blumenbach Institute of Zoology and Anthropology*; Georg-August-Universität Göttingen: Göttingen, Germany, 2013.
28. Fasemore, M.D.; Bidmos, M.A.; Mokoena, P.; Imam, A.; Billings, B.K.; Mazengenya, P. Dimensions around the Nutrient Foramina of the Tibia and Fibula in the Estimation of Sex. *Forensic Sci. Int.* **2018**, *287*, 222.e1–222.e7. [CrossRef]
29. Dayal, M.R.; Kegley, A.D.; Štrkalj, G.; Bidmos, M.A.; Kuykendall, K.L. The History and Composition of the Raymond, A. Dart Collection of Human Skeletons at the University of the Witwatersrand, Johannesburg, South Africa. *Am. J. Phys. Anthropol. Off. Publ. Am. Assoc. Phys. Anthropol.* **2009**, *140*, 324–335.
30. Curate, F. The Estimation of Sex of Human Skeletal Remains in the Portuguese Identified Collections: History and Prospects. *Forensic Sci.* **2022**, *2*, 272–286. [CrossRef]
31. Spradley, M.K.; Jantz, R.L. Sex Estimation in Forensic Anthropology: Skull Versus Postcranial Elements. *J. Forensic Sci.* **2011**, *56*, 289–296. [CrossRef]
32. Bruzek, J. A Method for Visual Determination of Sex, Using the Human Hip Bone. *Am. J. Phys. Anthropol.* **2002**, *117*, 157–168. [CrossRef]
33. Buikstra, J.E.; Ubelaker, D.H. Standards for Data Collection from Human Skeletal Remains. *Ark. Archaeol. Surv. Res. Ser.* **1994**, *44*, 10004710139.
34. Ferembach, D. Recomendations for Age and Sex Diagnosis of Skeletons. *J. Hum. Evol.* **1980**, *9*, 517–549.
35. Grupe, G.; Harbeck, M.; McGlynn, G.C. *Prehistoric Anthropology, Original Text*; Springer Spektrum: Berlin, Germany, 2015. [CrossRef]
36. Herrmann, B.; Grupe, G.; Hummel, S.; Piepenbrink, H.; Schutkowski, H. *Prehistoric Anthropology, Original Text: Guide to Field and Laboratory Methods*; Springer: Berlin, Germany, 1990.
37. Klepinger, L.L. *Fundamentals of Forensic Anthropology*; John Wiley & Sons: Hoboken, NJ, USA, 2006; Volume 1.
38. Novotný, V. Pohlavnı Rozdıly a Identifikace Pohlavá Pınevnı Kosti [Sex Differences and Identification of Sex in Pelvic Bone]. Ph.D. Thesis, Purkyne University, Ústí nad Labem, Czech Republic, 1981.
39. Loth, S.R.; Henneberg, M. Mandibular Ramus Flexure: A New Morphologic Indicator of Sexual Dimorphism in the Human Skeleton. *Am. J. Phys. Anthropol. Off. Publ. Am. Assoc. Phys. Anthropol.* **1996**, *99*, 473–485. [CrossRef]
40. McFadden, C.; Oxenham, M.F. Revisiting the P Henice Technique Sex Classification Results Reported by M Ac L Aughlin and B Ruce (1990). *Am. J. Phys. Anthropol.* **2016**, *159*, 182–183. [CrossRef]
41. Watson, P.F.; Petrie, A. Method Agreement Analysis: A Review of Correct Methodology. *Theriogenology* **2010**, *73*, 1167–1179. [CrossRef]
42. *Handbook of Forensic Anthropology and Archaeology*, 2nd ed.; Blau, S.; Ubelaker, D.H. (Eds.) Routledge: New York, NY, USA, 2016. [CrossRef]
43. Rowbotham, S.K. Anthropological Estimation of Sex. In *Handbook of Forensic Anthropology and Archaeology*; Routledge: New York, NY, USA, 2016; pp. 303–314.
44. Dunsworth, H.M. Expanding the Evolutionary Explanations for Sex Differences in the Human Skeleton. *Evol. Anthropol. Issues News Rev.* **2020**, *29*, 108–116. [CrossRef]
45. Krüger, G.C.; L'Abbé, E.N.; Stull, K.E.; Kenyhercz, M.W. Sexual Dimorphism in Cranial Morphology among Modern South Africans. *Int. J. Leg. Med.* **2015**, *129*, 869–875. [CrossRef]