

The effect of population variation on the accuracy of sex estimates derived from basal occipital discriminant functions

Sarah Inskip¹ · M. Constantinescu² · A. Brinkman¹ · M. Hoogland¹ · J. Sofaer³

Received: 3 September 2015 / Accepted: 29 August 2016 / Published online: 21 September 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Multiple discriminant functions that estimate sex from the dimensions of the basal occipital have been published. However, as there is limited exploration of basal dimension variation between groups, the accuracy of these functions when applied to archaeological material is unknown. This study compares basal dimensions between four known sex-at-death post-medieval European samples and explores how metric differences impact on the accuracy of sex assessment discriminant functions. Published data from St Bride's, London ($n = 146$) and the Georges Olivier collection, Paris ($n = 68$) were compared with new data from the eighteenth to nineteenth century Dutch Middenbeemster sample ($n = 74$) and the early twentieth century Rainer sample, Romania ($n = 282$) using independent t tests. The Middenbeemster and Rainer data were substituted into six published discriminant functions derived from the St Bride's and the Georges Olivier samples, and the results were compared to their known sex. Multiple statistically significant differences were found between the four groups. Of the six discriminant functions tested, five failed to reach the published accuracy and fell below chance. In addition, even where the samples were statistically comparable in means, trends for difference also impacted the accuracy of discriminant functions. Enough variation in basal occipital dimensions existed in the European

groups to decrease the accuracy of sex estimation discriminant functions to unusable. Possible inter-observer error, varying genetic, socioeconomic, and geographical factors are likely causes of dimension variation. This research further highlights the dangers of using sex estimation discriminant functions on samples that differ to the original derivative population and demonstrates the need for more rigorous testing.

Keywords Bioarchaeology · Sex assessment · Craniometrics · Multivariate analysis

Introduction

The use of discriminant functions to estimate sex has a long history in physical anthropology, and many equations have been created for different skeletal elements for use varying groups (for example see Bryers 2008 or Schwartz 2006). In the past few decades, a number of studies based on populations from across the globe have indicated the potential for using dimensions of the basal part of the occipital for the estimation of sex in fragmented human skeletal remains (Avci et al. 2010; Catalina-Herrera 1987; Franklin et al. 2013; Gapert et al. 2009a; Günay and Altinkök 2000; Holland 1986; Kajanoja 1969; Kanchan et al. 2013; Macaluso 2011; Manoel et al. 2009; Murshed et al. 2003; Naderi et al. 2005; Raghavendra Babu et al. 2012; Singh and Talwar 2013; Ukoha et al. 2011). The majority of these use the dimensions of the foramen magnum (width and length) to carry out discriminant function analyses and/or linear regression analysis to assess sex. These approaches have achieved sex assessment accuracy rates between 60 and 70 % for individual populations. Studies that use or include occipital condyle measurements have higher accuracy rates of up to 80 % (see Gapert et al. 2009b). Thus, while the occipital basal

✉ Sarah Inskip
s.a.inskip@arch.leidenuniv.nl

¹ Faculty of Archaeology, Leiden University, Leiden, South Holland 2333 CC, The Netherlands

² Laboratory of Paleoanthropology, "Francisc J. Rainer" Institute of Anthropology, Romanian Academy, 050474 Bucharest, Romania

³ Archaeology, Faculty of Humanities, University of Southampton, Avenue Campus, Southampton SO17 1BF, UK

measurements should not be used in isolation for sex estimation unless absolutely unavoidable (Gapert et al. 2009a; Wescott and Moore-Jansen 2001), the region does have sufficient sexual dimorphism to be of potential value when dealing with fragmented archaeological remains.

To date, however, there has been little comparative research exploring differences in basal dimensions between populations, and how significant any variation might be in terms of the accuracy of discriminant functions used for estimating sex. This is despite other researchers highlighting the accuracy problems of using discriminant functions on different groups when using other cranial measurements (Franklin et al. 2013; Kajanoja 1969). Although some researchers have suggested that no difference in basal dimensions exist between groups of differing biological ancestry (Holland 1986; Manoel et al. 2009; Naderi et al. 2005), others have voiced concerns over the accuracy of discriminant functions when applied to groups from dissimilar temporal contexts (Gapert et al. 2009b) and of differing biological ancestry (Wescott and Moore-Jansen 2001). Based on published mean dimensions, Gapert et al. (2009b) have already shown that the degree of sexual dimorphism for basal occipital dimensions varies between populations and argue that such differences could impact on sex assessment accuracy. It is not uncommon to see discriminant functions produced on one population being applied widely to disparate groups in terms of time or geography. While some skeletal dimensions appear stable enough in terms of size and sexual dimorphism between groups for wider use, others do not and require population-specific functions.

The aim of this paper is to test whether sex estimation discriminant functions based on basal occipital dimensions are accurate when applied to material from outside their original deriving collection. This will be achieved by first exploring general variation in the dimensions of the basal occipital region between four post-medieval European collections of known sex at death. Second, published discriminant functions derived from two of the collections will be tested using measurements from the two other collections. Overall, this allows us to assess whether sex assessment discriminant functions using basal occipital dimensions created on European samples can be used on other European groups of a similar date, or if there is not enough stability in basal measurements for the

discriminant functions to be used more widely, as has been the case with other dimensions.

Materials and methods

Data from four different skeletal collections were used to undertake this research. First, two collections with published discriminant functions were selected. This was the eighteenth–nineteenth century English urban sample from St Bride's, London, which represents a middle to high class group (Gapert et al. 2009a, b), and the twentieth century urban poor French sample from the Georges Olivier collection, Paris (Macaluso 2011). The selection criterion was based on the use of the same measurement methods and the availability of basal occipital dimension data, including means, standard deviation, and number of individuals, by sex. In addition, raw data for the Georges Olivier sample was obtained. For both collections, the three discriminant functions which produced the highest cross-validated accuracy were selected for testing (see Table 1).

To test the accuracy of these functions, two other known sex-at-death collections were required. The Rainer skeletal collection, which is housed at Institutul de Anthropologie 'Francisc J. Rainer' Bucharest, Romania, contains the remains of over 6000 individuals. These were collected from 33 local hospitals over a period of 50 years with the majority dating from the 1930s to 1940s. They represent an urban poor community (Ion 2011). Approximately 50 % of the individuals in the collection are of known identity, including age, sex, and ethnicity. The date, cause, and location of death are also recorded. A random sample of 282 adult individuals over 18 years of age with well-preserved and complete occipitals were selected and measured.

A second, smaller sample of individuals was selected from the Middenbeemster collection housed at the Laboratory for Human Osteoarchaeology and Funerary Archaeology at Leiden University, The Netherlands. This collection consists of over 450 working class individuals from a rural farming community dating from the late eighteenth century to the mid-nineteenth century. Approximately one quarter of the individuals are identifiable from archival records. Of these 74

Table 1 The discriminant functions from the St Bride's sample (Gapert et al. 2009a, b) and the Georges Olivier sample (Macaluso 2011) tested on the Rainer and Middenbeemster samples

Source	Abbreviation	Measurements
Gapert et al. 2009b	GF1	MLC right, MWC right, MLC left, MWC left, BCB, MxID, MnD
	GF2	MLC left, MWC right, MnD
	GF3	BCB and MnD
Macaluso 2011	MF1	MLC left and MnD
	MF2	MLC left, MLC right, MWC left, MWC right
	MF3	BCB

individuals had crania complete enough to be included in the study. Table 2 presents the descriptive statistics for sex and age for the Rainer and Middenbeemster samples.

Collection, analysis, and comparison of basal occipital dimensions

The occipital condyle and foramen magnum measurements used followed Gapert et al. (2009a, b) and are based on Holland (1986) and Wescott and Moore-Jansen (2001). These are depicted in Figs. 1, 2, and 3. The measurements are outlined in Table 3; the external hypoglossal canal distance measurement was not taken for the Middenbeemster sample. All measurements were taken twice to permit an analysis of intra-observer error. In addition, to further minimize error, when there was deviation (over 0.5 mm) between the first and second measurement, the dimension was remeasured the third time and the two closest measurements were used. For all other statistical testing, the average of the two measurements was used.

Intra-observer error of repeatability was tested on the Middenbeemster and Rainer collection. The absolute technical error of measurement (TEM), relative technical error of measurement (rTEM), and coefficient of reliability (R) were calculated following Perini et al. (2005) and Gapert et al. (2009b) to assess the degree of magnitude of the random error of measurement. Inter-observer error rates for the measurements used can be found in Gapert and Last (2005) and Wescott and Moore-Jansen (2001).

Prior to any inter-site comparison of basal occipital dimensions, the Middenbeemster and Rainer data were tested for normality using a Kolmogorov-Smirnov test and Levene's test for equality of variance. Box's M tests were used to test for equality of covariances. In order to see if significant differences in occipital dimensions existed between the sexes of the Rainer and Middenbeemster groups, two-tailed independent sample t tests were carried out.

All means were compared between all groups. Without the raw data from the St Bride's sample, it was not possible to carry out ANOVA tests. While this would have been preferable, it was possible to undertake independent t tests (two tailed) to compare the means between groups with the sexes

being analyzed separately. Bootstrapping of 1000 samples was carried out on the comparisons between the Rainer, Georges Olivier, and Middenbeemster samples to control for the unequal sample sizes. As age and head/body size have been shown to not correlate with basal dimensions (Gapert et al. 2013; Guidotti 1984; Naderi et al. 2005; Wescott and Moore-Jansen 2001), they were not tested in the present paper.

After mean comparisons, the measurements from the Middenbeemster and Rainer individuals were substituted into the discriminant functions derived from Gapert et al. (2009b) and Macaluso (2011) to create discriminant scores. The sectioning point associated with the equation was then used to classify individuals as male or female. A score above the sectioning point represented male, and a score below represented female. This classification was then compared to the known sex of the individual. Statistical significance was set at $p < 0.05$. All testing were completed in SPSS 21.0.

Results

Table 4 presents the results of the intra-observer tests including the absolute technical error of measurement (TEM), relative technical error of measurement (rTEM), and coefficient of reliability (R). The scores show that there was good accuracy for each measurement for the Rainer and Middenbeemster samples. The high values for R indicate that the variance was unlikely to be caused by human error. In both samples, the measurement with the least variation was the bicondylar breadth. The measurements with the greatest variation were the width of the left and right condyles. These were also the two measurements with the lowest R values.

According to the Kolmogorov-Smirnov tests, all variables were normally distributed ($p > 0.05$). Levene's tests demonstrated equality of variances ($p > 0.05$) with the exception of the right maximum width of the condyle in the Middenbeemster sample ($p = 0.021$, $n = 23$). A similar result was experienced by Gapert et al. (2009b).

Table 5 outlines the descriptive statistics for the dimensions of the occipital bone as well as the results of independent t tests for sex differences for the 282 Rainer individuals. Table 6 contains the same data for the individuals from Middenbeemster,

Table 2 Descriptive age and sex statistics for the Rainer (Romanian) and Middenbeemster (Dutch) samples

Sample	Sex	Mean age	Min age	Max age	Range	SD	N
Rainer	Female	39.5	18	88	70	15.98	147
	Male	45.0	18	86	68	15.25	135
Middenbeemster	Female	48.8	23	84	62	16.77	42
	Male	55.9	22	84	61	20.47	32

SD standard deviation, N number of individuals

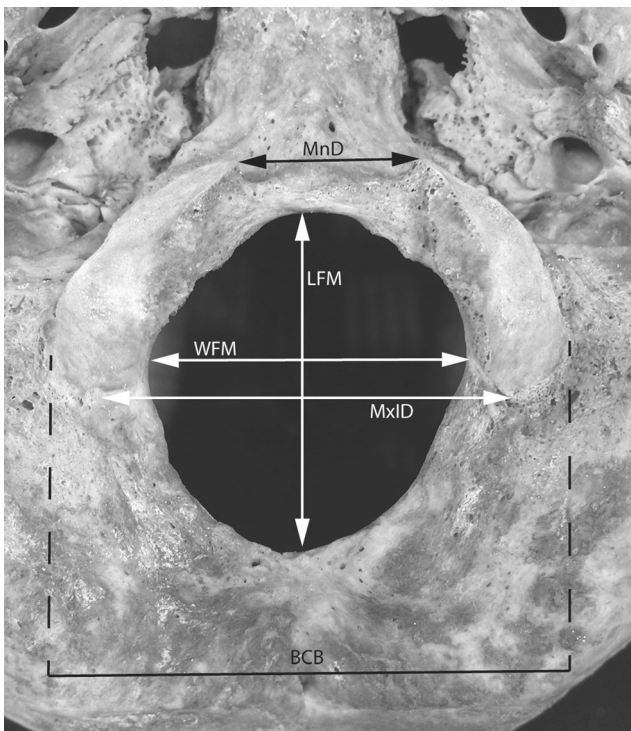


Fig. 1 Basal occipital measurements used in this study. *BCB* bicondylar breadth, *MxID* maximum intercondylar distance, *MnD* minimum intercondylar distance, *LFM* length of the foramen magnum, *WFM* width of the foramen magnum

with the exception of the right maximum width of the condyle which displays the *p* value for equal variances not assumed.

For the Rainer and Middenbeemster samples, all male dimensions were larger than those of females. In the Rainer collection, the differences between males and females were statistically significant with the exception of the maximum intercondylar distance which was just outside significance (Table 5). With the exception of the maximum length of the left condyle, the maximum width of the right and left condyle,

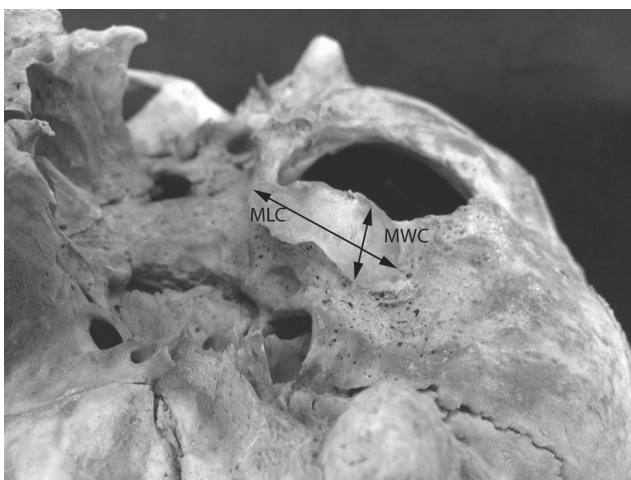


Fig. 2 Occipital condyle measurements. *MLC* maximum length of occipital condyle, *MWC* maximum width of the occipital condyle

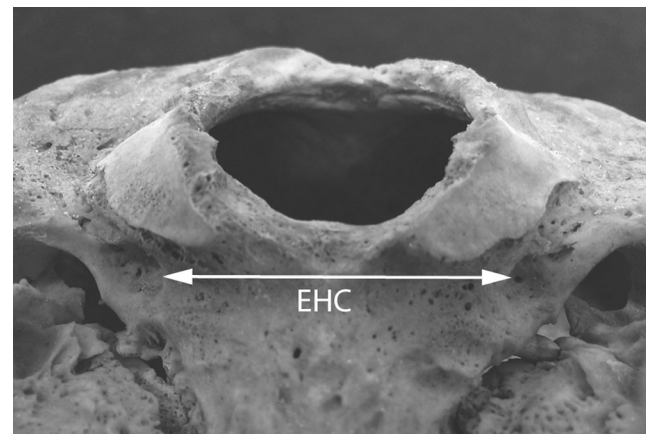


Fig. 3 Depiction of the measurement of the distance between the external hypoglossal canals

and the maximum intercondylar distance, the remaining dimensions in Middenbeemster were also statistically significantly different between the sexes (Table 6).

The means, standard deviations, and number of individuals for the Georges Olivier and St Bride's samples are presented in Table 7. The results of two-tailed independent *t* tests comparing the means between the four groups are presented in Table 8.

The *t* test results show that there are 23 statistically significant differences between the samples. The most variable dimension was the minimum intercondylar distance followed by the external hypoglossal canal distance and the occipital condyle length. The least variable dimensions were the maximum widths of the occipital condyles, which showed no statistical differences between any of the groups tested. The measurements do not indicate a clear trend for one sample to have all of the largest or the smallest dimensions with the largest and smallest dimensions for each variable being distributed between different groups. This may suggest variation in the metric relationship between the different dimensions of the basal occipital region between groups. Of the four groups, the Rainer sample had the most statistically significant differences with the other three samples; there are 11 statistically

Table 3 Basal occipital measurements and their abbreviations taken on the Middenbeemster and Rainer skulls

Measurement and abbreviation	Source
Maximum length of occipital condyle (MLC)	Gapert et al. 2009b
Maximum width of occipital condyle (MWC)	Gapert et al. 2009b
Bicondylar breadth (BCB)	Gapert et al. 2009b
Maximum intercondylar width (MxID)	Gapert et al. 2009b
Minimum condylar width (MnD)	Gapert et al. 2009b
External hypoglossal canal distance (EHC)	Gapert et al. 2009b
Length of foramen magnum (LFM)	Gapert et al. 2009a
Width of foramen magnum (WFM)	Gapert et al. 2009a

Table 4 The results of intra-observer error tests with TEM, rTEM, and *R* for the Rainer and Middenbeemster samples

Measurement	Rainer			Middenbeemster		
	Absolute TEM (mm)	Relative TEM (%)	<i>R</i>	Absolute TEM (mm)	Relative TEM (%)	<i>R</i>
MLC left	0.13	0.55	0.9975	0.18	0.9963	0.9963
MLC right	0.12	0.94	0.9972	0.19	0.9958	0.9958
MWC left	0.13	1.11	0.9917	0.15	0.9850	0.9850
MWC right	0.12	1.00	0.9921	0.16	0.9871	0.9871
EHC	0.09	0.29	0.9997	–	–	–
BCB	0.13	0.25	0.9985	0.12	0.9989	0.9989
MxID	0.17	0.33	0.9985	0.21	0.9968	0.9968
MnD	0.12	0.59	0.9980	0.17	0.9943	0.9943
LFM	0.12	0.34	0.9978	0.10	0.9983	0.9983
WFM	0.11	0.37	0.9976	0.13	0.9972	0.9972

The intra-observer error testing was carried out on the original sample of 311 Rainer skulls prior to elimination of pathological specimens from the study

significant differences between the Rainer and St Bride’s samples, four with the Georges Olivier sample, and five with the Middenbeemster sample (Table 8).

When the Rainer and Middenbeemster measurements were substituted into the St. Bride’s discriminant functions published in Gapert et al. (2009b), all three performed poorly (Table 9). For the first function (GF1), no females were correctly identified and nearly all individuals were sexed as male. A similar trend was observed for GF3 where few females were

correctly identified. GF2 produced similar accuracy rates for male identification as the original sample, but again there was very poor accuracy when it came to identifying females.

Macaluso’s functions based on the Georges Olivier material performed better on the Rainer sample than the St Bride’s functions (Table 10). Macaluso’s (2011) stepwise function (MF1) obtained similar sex assessment accuracy as the original study when applied to the Rainer sample, but there was a 7.2 % increase in sex bias towards males. While MF2 had

Table 5 Descriptive statistics for the 282 skulls of the Rainer sample and results of two-tailed independent *t* tests for sex differences

Measurement	Sex	Mean	S.E	SD	Min	Max	<i>N</i>	<i>t</i> test	<i>p</i>
MLC left	♀	22.63	0.25	2.49	11.28	27.74	147	6.20	<0.001
	♂	24.33	0.30	2.34	19.18	28.44	135		
MLC right	♀	22.07	0.22	2.21	14.10	27.96	147	5.88	<0.001
	♂	24.12	0.26	2.04	18.83	30.16	135		
MWC left	♀	11.61	0.14	1.45	8.43	15.90	147	3.18	0.002
	♂	12.02	0.15	1.15	9.31	15.20	135		
MWC right	♀	11.59	0.13	1.33	8.97	15.70	147	3.68	<0.001
	♂	11.92	0.15	1.15	9.75	15.10	135		
BCB ave	♀	49.60	0.28	2.87	42.78	57.55	147	5.21	<0.001
	♂	51.91	0.41	3.22	43.96	59.69	135		
MnD ave	♀	19.85	0.24	2.42	14.28	26.77	147	2.26	0.025
	♂	20.44	0.36	2.86	14.62	28.74	135		
MxID ave	♀	41.34	0.31	3.17	33.28	48.96	147	1.83	0.068
	♂	42.17	0.44	3.47	35.38	51.56	135		
Df ave	♀	40.38	0.34	3.45	33.17	50.01	147	3.15	0.002
	♂	42.22	0.50	3.90	33.15	52.23	135		
EHC ave	♀	32.19	0.20	2.05	26.76	38.12	147	8.87	<0.001
	♂	34.73	0.27	2.14	30.14	40.77	135		
LFM ave	♀	34.49	0.23	2.35	26.61	41.25	147	6.13	<0.001
	♂	36.24	0.31	2.44	29.42	43.25	135		
WFM ave	♀	29.61	0.18	1.81	25.83	33.15	147	6.66	<0.001
	♂	31.21	0.19	2.22	25.54	37.72	135		

The italics are statistically significant at the *p*=0.05 level

SE standard error, *SD* standard deviation, *min* minimum, *max* maximum, *N* number

Table 6 Descriptive statistics for the 74 Middenbeemster skulls and results of two-tailed independent *t* tests for sex differences

Measurement	Sex	Mean	SE	SD	Min	Max	<i>N</i>	<i>t</i> test	<i>p</i>
MLC left	♀	24.31	0.48	2.61	19.87	28.61	30	-1.407	0.166
	♂	25.47	0.71	3.32	18.95	30.24	22		
MLC right	♀	23.79	0.48	2.52	19.17	28.47	35	-2.137	0.037
	♂	24.97	0.74	3.30	16.51	32.66	22		
MWC left	♀	11.76	0.22	1.13	9.59	14.44	29	-0.883	0.382
	♂	12.06	0.31	1.39	8.39	14.36	20		
MWC right	♀	11.25	0.23	1.20	9.04	13.09	35	-1.709	0.096
	♂	11.83	0.36	1.62	9.72	14.96	22		
BCB	♀	49.82	0.67	3.47	43.27	57.93	28	-2.58	0.013
	♂	52.31	0.74	3.31	48.56	61.83	20		
MnD	♀	19.30	0.38	1.98	15.11	24.11	30	-3.01	0.004
	♂	20.82	0.53	2.36	17.20	26.30	22		
MxID	♀	35.13	0.64	3.34	29.37	41.77	31	-1.413	0.164
	♂	36.11	0.97	4.32	25.72	46.46	21		
LFM	♀	34.96	0.49	2.56	30.36	40.47	38	-2.493	0.015
	♂	36.43	0.47	2.10	32.73	39.47	28		
WFM	♀	29.47	0.46	2.38	25.02	34.41	38	-2.983	0.004
	♂	31.06	0.52	2.32	27.65	36.61	28		

The italics are statistically significant at the $p=0.05$ level

SE standard error, SD standard deviation, min minimum, max maximum, N number

higher sex pooled accuracy rates when used on the Rainer collection, there was increased sex bias towards females,

Table 7 Published means, standard deviations, and number of individuals for the Georges Olivier and the St Brides samples

Measurement	Sex	St Bride's ^a			Georges Olivier ^b		
		Mean	SD	<i>N</i>	Mean	SD	<i>N</i>
MLC left	♀	23.74	2.44	71	22.88	2.69	31
	♂	25.16	2.51	75	24.99	3.09	35
MLC right	♀	23.30	2.28	71	22.99	2.28	31
	♂	24.95	2.53	75	24.62	2.65	35
MWC left	♀	11.57	1.16	71	11.57	1.09	31
	♂	12.05	1.69	75	12.25	1.51	35
MWC right	♀	11.42	1.21	71	11.59	1.03	31
	♂	12.01	1.41	75	12.30	1.27	35
BCB	♀	48.67	3.17	71	48.73	3.27	32
	♂	51.29	2.97	75	51.32	3.70	36
MnD	♀	19.00	2.40	71	19.07	2.14	31
	♂	21.12	3.18	75	20.63	3.18	35
MxID	♀	35.12	3.09	71	36.78	3.69	32
	♂	36.82	3.10	75	37.46	3.54	36
EHC	♀	31.60	2.12	71	33.29	1.88	32
	♂	33.37	2.35	75	34.00	2.30	36
LFM	♀	34.78	1.97	71	34.90	2.26	32
	♂	35.79	2.36	75	35.38	2.27	36
WFM	♀	29.35	2.06	71	29.40	2.63	32
	♂	30.48	1.86	75	30.72	2.11	36

^a St Bride's see Gapert et al. 2009a,b

^b Georges Olivier see Macaluso 2011

which likely resulted from the larger female condyles in the Georges Olivier sample. MF3 produced a large sex bias in favor of male identification. This is because the mean bicondylar breadth was statistically larger in the Rainer group putting many of the females over the sectioning point and the males at the extreme end formula.

When applied to the Middenbeemster collection, MF1 had a 17.4 % increase in sex bias towards males, so although more males were successfully identified, fewer females were correctly identified effectively decreasing the overall sex pooled accuracy. Conversely, MF2 and MF3 produced similar accuracy rates to those obtained on the original deriving sample. For MF2, there was a 6.4 % decrease in the number of correctly identified males and a 4.5 % increase in correctly identified females. This reduced the original sex bias to just 0.5 %. A similar phenomenon was seen for the third function where the sex bias was also reduced from -13.9 % to 5.7 %. With the reduction in sex bias, it thus appears that MF2 and MF3 actually performed better on the Middenbeemster group than on the original sample.

Discussion

The accuracy of discriminant functions used to estimate sex relies on similarity between the individual/group being tested and the sample population on which the function was originally derived. While some authors have presented dimensions from multiple groups (Gapert et al. 2009a, b; Macaluso 2011; Ukoha et al. 2011), to date there has been little statistical analysis of differences between groups, and no previous

Table 8 Results of two-tailed independent *t* tests comparing basal dimensions between the four collections by sex

Measurement	Sex	Rainer vs. St Bride's		Rainer vs. Middenbeemster ^a		Rainer vs. Georges Olivier ^a		Middenbeemster vs. St Bride's		Middenbeemster vs. Georges Olivier ^a		Georges Olivier vs. St Bride's	
		<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>	<i>p</i>	<i>t</i>
MLC left	♀	<i>0.002</i>	<i>31.630</i>	<i>0.011</i>	<i>2.980</i>	0.754	0.340	0.292	10.601	<i>0.025</i>	<i>2.181</i>	0.116	15.870
	♂	<i>0.018</i>	<i>23.960</i>	0.207	1.613	0.557	-0.703	0.641	0.467	0.496	0.716	0.760	0.307
MLC right	♀	<i><0.001</i>	<i>38.800</i>	<i>0.019</i>	<i>2.358</i>	0.106	-1.495	0.312	10.150	0.302	1.067	0.529	0.632
	♂	<i>0.012</i>	<i>25.456</i>	0.203	1.661	0.225	-1.352	0.976	0.030	0.674	0.435	0.531	0.628
MWC left	♀	0.874	0.155	0.308	0.921	0.906	-0.109	0.452	0.754	0.440	0.782	1.000	0.000
	♂	0.881	0.150	0.889	0.127	0.389	-0.845	0.981	0.024	0.649	-0.450	0.551	0.598
MWC right	♀	0.357	0.923	0.586	-0.571	0.494	-0.623	0.673	0.423	0.255	-1.163	0.497	0.681
	♂	0.623	0.492	0.610	-0.564	0.234	-1.134	0.605	0.519	0.302	-1.204	0.302	1.036
BCB	♀	<i>0.029</i>	<i>22.035</i>	0.881	0.165	0.186	1.488	0.114	15.938	0.237	1.136	0.930	0.088
	♂	0.179	13.468	0.348	0.983	0.720	0.389	0.190	13.193	0.308	1.004	0.965	0.0439
MnD	♀	<i>0.014</i>	<i>24.821</i>	<i>0.052</i>	<i>-1.719</i>	0.076	16.530	0.552	0.600	0.691	0.365	0.900	0.126
	♂	0.121	15.575	0.730	0.287	0.972	-0.023	0.684	0.408	0.803	0.231	0.453	0.753
MxID	♀	<i><0.001</i>	<i>139.449</i>	<i>0.001</i>	<i>-10.590</i>	<i>0.001</i>	<i>7.078</i>	0.988	0.015	0.104	-1.733	<i>0.020</i>	<i>23.727</i>
	♂	<i><0.001</i>	<i>108.892</i>	<i>0.001</i>	<i>0.315</i>	<i>0.001</i>	<i>7.558</i>	0.401	0.843	0.286	-1.179	0.333	0.972
EHC	♀	<i>0.049</i>	<i>19.750</i>			<i>0.003</i>	<i>-2.785</i>	-	-			<i><0.001</i>	<i>38.731</i>
	♂	<i><0.001</i>	<i>41.835</i>			0.221	1.201	-	-			0.190	13.312
LFM	♀	0.360	0.9167	0.565	0.586	0.574	-0.539	0.678	0.414	0.858	0.174	0.785	0.273
	♂	0.205	12.706	0.850	0.655	<i>0.032</i>	<i>2.114</i>	0.212	12.574	0.078	1.707	0.388	0.867
WFM	♀	0.335	0.965	0.603	-0.521	0.815	0.312	0.782	0.277	0.899	-0.119	0.467	0.730
	♂	<i>0.017</i>	<i>24.149</i>	0.783	0.459	0.207	1.259	0.193	13.098	0.548	0.609	0.544	0.609

The italics are statistically significant at the *p*=0.05 level

^a Bootstrapped analysis

studies have tested the accuracy of discriminant functions on other known populations. The results show that despite the suggestions of some researchers that there is little difference in dimensions between groups of differing biological ancestry (Holland 1986; Manoel et al. 2009), there is enough variation between the European groups tested here to significantly impact the accuracy of sex assessment discriminant functions. This supports research by other scholars who have indicated that there may be differences between groups (Gapert et al. 2009b; Wescott and Moore-Jansen 2001). Thus, while it seems that the sex assessment accuracy obtained from basal

occipital discriminant functions is generally similar between studies, ranging from 60 to 80 %, the application of a single set of discriminant functions on diverse populations of European ancestry is highly problematic.

When using the Gapert functions, a significant decrease in accuracy from that obtained on the original sample was observed. In particular, it was clear that the larger means of females in the Rainer group led to them being classified as males. In addition, as the dimensions of Rainer males were also larger, their discriminant scores were well over the sectioning point resulting a significant sex bias. A similar

Table 9 Percent (%) correct sex estimations from the original Gapert et al. (2009a, b) functions and the % correct predictions of sex predictions for the Rainer and Middenbeemster individuals (GF1 = MLC left, MLC right, MWC left, MWC right, BCB, MnD, and MxID; GF2 = MLC left, MWC right, and MnD; GF3 = BCB and MnD)

Sample	Female %	Males %	Pooled sex %	<i>N</i>	Sex bias
GF1 original test result	76.1	72.0	74.0	146	-4.1
Rainer	0.0	100.0	47.9	282	100.0
Middenbeemster	0.0	100.0	41.7	48	100.0
GF2 original test result	81.7	72.0	76.7	146	-9.7
Rainer	51.7	76.3	63.5	282	24.6
Middenbeemster	44.8	76.2	58.0	50	31.4
GF3 original test result	74.6	73.3	74.0	146	-1.3
Rainer	17.0	91.8	52.8	282	74.8
Middenbeemster	21.4	100	54.2	48	78.6

Table 10 Percent (%) correct sex predictions from the original Macaluso (2011) functions and the % correct predictions of sex predictions for the Rainer and Middenbeemster individuals (MF1 = MLC left and MnD; MF2: MLC left, MLC right, MWC left, and MWC right; MF3: BCB)

Sample	Female %	Males %	Pooled sex %	<i>N</i>	Sex bias
MF1 original test result	66.7	68.6	67.7	65	1.9
Rainer	61.9	71.0	66.3	282	9.1
Middenbeemster	51.7	71.0	60.0	50	19.3
MF2 original test result	60.0	71.4	66.2	65	11.4
Rainer	75.5	63.7	69.9	282	-11.8
Middenbeemster	64.5	65.0	64.7	51	0.5
MF3 original test result	75.0	61.1	67.6	68	-13.9
Rainer	54.4	70.4	62.0	282	16.0
Middenbeemster	64.3	70.0	66.7	48	5.7

problem was incurred for the Middenbeemster sample which was even more interesting considering the lack of statistically significant differences with the St Bride's group. The failure of the functions can be explained by the process of creating discriminant scores in which multiple values obtained from different dimensions are added together. Although there were no individual dimensions that were statistically significantly larger in the Middenbeemster group, so the samples appear metrically comparable, the fact that the Middenbeemster individuals, and the females in particular, were slightly larger for multiple dimensions resulted in a higher discriminant score which placed many of the women over Gapert et al.'s (2009b) sectioning point.

While when testing the first and third functions of Macaluso (2011) on the Rainer sample, a similar misclassification of females was observed; the opposite problem could be seen for the second function. This function relies solely on the size of the occipital condyles which were smaller in the Rainer collection. This meant that the scores for males were below the sectioning point, while the female scores were all well below the sectioning point. This means that more females were correctly assessed and more males were misclassified. When applied to the Middenbeemster sample, the second and third functions from Macaluso (2011) were the only functions that produced similar results to the original tests. In fact, the sex biases identified in the original testing were reduced in the Middenbeemster group as a result of the increased size of Middenbeemster individuals. It was unclear why this occurred because the standard deviations appeared similar between the two samples and the sexual dimorphism was lower in the Middenbeemster group. As such, further comparison between these two samples is required, although the small size of the function's deriving sample cannot be ruled out. Overall, the sectioning points created on the St Bride's and Georges Olivier material were generally not suitable for the Rainer and Middenbeemster collections.

Before speculating on possible reasons as to why the cranial measurements differed between groups and the discriminant functions failed, it is important to discuss inter-observer

error as it acts as a source of metric variation. As this paper used published data, it was not possible to undertake inter-observer error tests, and it is possible that this may play a role in the differences between the samples. Research has shown that with the exception of the occipital condyle width and length, the dimensions of the basal occipital assessed here can be well replicated (Gapert and Last 2005; Wescott and Moore-Jansen 2001) which suggests that inter-observer error can be limited with careful measurement. Here, every care was taken to minimize this error which is reflected in the intra-observer tests results. It is perhaps significant that even the most replicable dimensions (bicondylar breadth, external hypoglossal canal, and the width of the foramen magnum) also produced statistically significant differences.

There are a number of possible reasons why basal occipital measurements differed between the groups analyzed here. While they could be broadly classed as "European," they still differ in their genetic, social economic, and environmental conditions. It is of course interesting to note that it was the Rainer sample that was most different to the other samples. This is perhaps unsurprising given that it is by far the most geographically distant of the collections assessed here. However, differences may also arise from general increases in overall size, as has been suggested by Gapert et al. (2009b), but also variation in nutrition and disease load, both of which impact on the growing skeleton and final adult size (Larsen 2015). This makes it difficult to say exactly what the sources of disparity were in this case. As such, more dedicated intra- and inter-population analysis of basal occipital dimensions in known samples that control for these factors are required. In addition, it would be beneficial to obtain the raw data for all of the samples and undertake more complex statistical testing on the data. This includes M/ANOVA testing, multiple discriminant analysis, or principle component analysis. Until these factors are analyzed in greater detail and we have a better understanding of their relationship with the various basal occipital dimensions, then it is unwise to apply basal occipital functions to collections other than its deriving group.

Conclusion

The aim of this research was to test whether discriminant functions based on basal occipital measurements could accurately estimate sex in other populations. It demonstrated that significant variation existed in the dimensions of the basal occipital between the four different post-medieval European groups and that these differences dramatically decreased the accuracy of the discriminant functions derived from two of the samples. Critically, even in groups that statistically have similar mean dimensions as the function deriving sample, the accuracy can decrease even if there are trends for difference. Many factors may be responsible for the differences between the groups, but until more is known about the relationship between these and basal occipital dimensions, it is problematic to apply basal occipital discriminant functions to any group other than the original deriving population.

Acknowledgments Many thanks to the reviewers for their time and helpful comments on this paper. Thanks are also due to James Macaluso for access to his unpublished data.

This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS-UEFISCDI, project number PNII-ID-PCCE-2011-2-0013.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Avcı E, Dagtekin A, Ozturk AH, Kara E, Ozturk NC, Uluc K, Aktüre E, Baskaya MK (2010) Anatomical variations of the foramen magnum, occipital condyle and jugular tubercle. *Turkish Neurosurgery* 21: 181–190
- Bryers S (2008) *Introduction to forensic anthropology*. Pearson Education, Boston
- Catalina-Herrera CJ (1987) Study of the anatomic metric values of the foramen magnum and its relation to sex. *Acta Anat* 130:344–347
- Franklin D, Cardini A, Flavel A, Kuliukas A (2013) Estimation of sex from cranial measurements in a Western Australian population. *Forensic Sci Int* 229:158.e1–158.e8
- Gapert R, Last J (2005) The adult human cranial bone: measurement variance and observer error. In: Zakrzewski SR, Clegg M (eds) *Proceedings of the Fifth Annual Conference of the British Association for Biological Anthropology and Osteoarchaeology*, BAR International Series. Archaeopress, Oxford, pp. 119–122
- Gapert R, Black S, Last J (2009a) Sex determination from the foramen magnum: discriminant function analysis in an eighteenth and nineteenth century British sample. *Int J Legal Med* 123:25–33
- Gapert R, Black S, Last J (2009b) Sex determination from the occipital condyle: discriminant function analysis in an eighteenth and nineteenth century British sample. *Am J Phys Anthropol* 138:384–394
- Gapert R, Black S, Last J (2013) Test of age-related variation in the craniometry of the adult human foramen magnum region: implications for sex determination methods. *Forensic Science, Medicine and Pathology* 9:478–488
- Guidotti A (1984) Morphometrical considerations on occipital condyles. *Anthropol Anz* 42:117–119
- Günay Y, Altinkök M (2000) The value of the size of foramen magnum in sex determination. *Journal of Clinical Forensic Medicine* 7:147e9
- Holland TD (1986) Sex determination of fragmentary crania by analysis of the cranial base. *Am J Phys Anthropol* 70:203–208
- Ion A (2011) A brief overview of “Francisc J. Rainer” human osteological collection. *Annals of Romanian Anthropology* 48:24–32
- Kajanoja P (1969) Sex determination of Finnish crania by discriminant function analysis. *Arch Oral Biol* 24:29–34
- Kanchan T, Gupta A, Krishan K (2013) Craniometric analysis of foramen magnum for estimation of sex. *International Journal of Medical, Pharmaceutical Science and Engineering* 7:97–99
- Larsen CS (2015) *Bioarchaeology. Interpreting behaviour from the human skeleton*. Cambridge University Press, Cambridge
- Macaluso PJ Jr (2011) Metric sex determination from basal region of the occipital bone in a documented French sample. *Bulletins et Mémoires de la Société d’anthropologie de Paris* 23:19–26
- Manoel C, Prado FB, Caria PHF, Groppo FC (2009) Morphometric analysis of the foramen magnum in human skulls of Brazilian individuals: its relation to gender. *Brazilian Journal of Morphological Sciences* 26:104–108
- Murshed KA, Çiçekcibaşı AE, Tuncer I (2003) Morphometric evaluation of the foramen magnum and variations in its shape: a study on computerized tomographic images of normal adults. *Turkish Journal of Medical Sciences* 33:301–306
- Naderi S, Korman E, Çitak G, Güvençer M, Arman C, Şenoğlu TS, Arda MN (2005) Morphometric analysis of human occipital condyle. *Clin Neurol Neurosurg* 107:191–199
- Perini TA, Oliveria G, Omellas J, Oliveria F (2005) Technical error of measurement in anthropometry. *Rev Bras Med Esporte* 11:86–90
- Raghavendra Babu YP, Kanchan T, Attiku Y, Dixit PN, Kotian MS (2012) Sex estimation from foramen magnum dimensions in an Indian population. *Int J Legal Med* 19:162–167
- Schwartz JH (2006) *Skeleton keys. An introduction to human skeletal morphology, development, and analysis*. Oxford University Press, Oxford
- Singh G, Talwar I (2013) Morphometric analysis of foramen magnum in human skull for sex determination. *Human Biology Review* 2:29–41
- Ukoha U, Egwu OA, Okafor IJ, Anyabolu AE, Ndukwe GU, Okpala I (2011) Sexual dimorphism in the foramen magnum of Nigerian adult. *International Journal of Biological & Medical Research* 2: 878–881
- Wescott D, Moore-Jansen P (2001) Metric variation in the human occipital bone: forensic anthropological applications. *Journal of Forensic Science* 5:1159–1163