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

Inskip S1, Constantinescu M2, Brinkman A1, Hoogland M1 and Sofaer J3.

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, Bucharest 050474, Romania

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

Number of text pages 13, Number of figures 3, Number of tables 11, Number of charts 0.

Abbreviated title: Occipital discriminant functions in European groups

Key words: Bioarchaeology, Sex assessment, craniometrics, multivariate analysis

Corresponding author: Sarah Inskip

s.a.inskip@arch.leidenuniv.nl

Faculty of Archaeology

Leiden University

2333 CC Leiden

The Netherlands.

Acknowledgements

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.

**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 18th-19th century Dutch Middenbeemster sample (n=74) and the early 20th century Rainer sample, Romania (n=282) using independent *t-tests.* The Middenbeemster and Rainer data was 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 in addition to 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.

**INTRODUCTION**

The use of discriminant functions to estimate sex has a long history in physical anthropology, and many equations for use on differing skeletal elements and varying groups exist (for example see Schwartz 2006). In the past few decades, a number of studies focusing 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 60-70% 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 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 the 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 18th - 19th century English urban sample from St Bride’s, London, which represents a middle to high class group (Gapert et al*.* 2009a, b), and the 20th century urban poor French sample from the Georges Olivier collection, Paris (Macaluso 2011). The selection criterion were 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).

[Table 1 here].

To test the accuracy of these functions, two other known sex-at-death collections were required. The Rainer skeletal collection is housed at Institutul de Anthropologie ‘Francisc J. Rainer’ Bucharest, Romania. The remains of over 6000 individuals were collected from 33 local hospitals over a period of 50 years, the majority dating from the 1930’s and 1940’s, and 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 18th century to the mid-19th century. Approximately one quarter of the individuals are identifiable from archival records. Of these 74 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.

[Table 2 here].

 **Collection, analysis and comparison of basal occipital dimensions**

The occipital condyle and foramen magnum measurements used followed those of Gapert et al*.* (2009a, b) and are based on Holland (1986) and Wescott and Moore-Jansen (2001). These are depicted in figures 1 -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 minimise error, when there was deviation (over 0.5mm) between the first and second measurement the dimension was remeasured a third time and the two closest measurements were used. For all other statistical testing the average of the two measurements was used.

[Table 3 here].

[Figure 1 here].

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

[Figure 2 here].

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

[Figure 3 here].

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

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 samples *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 was completed in SPSS 21.0.

**RESULTS**

Tables 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.

[Table 4 here].

According to the Kolmogorov Snirnov 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, with the exception of the right maximum width of the condyle of the iled)suggests different relaionships between certain features in different groupsdifferences in skull shape or size exis which displays the p value for equal variances not assumed.

[Table 5 here].

[Table 6 here].

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 and 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.

[Table 7 here].

[Table 8 here].

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 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.

[Table 9 here].

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

[Table 10 here].

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 use of discriminant functions to estimate sex relies on there being similarity between the individual 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*.* 2009; Macaluso 2011; Ukoha et al*.* 2011), to date there has been little statistical analysis of differences between groups, and no previous 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 60-80%, 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 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 Gapertet 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 sample 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 the 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 dramatically 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.

ACKNOWLEDGEMENTS

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.

**LITERATURE CITED**

Avci E, Dagtekin A, Ozturk AH, Kara E, Ozturk NC, Uluc K, Akture E, Baskaya MK (2010) Anatomical Variations of the Foramen Magnum, Occipital Condyle and Jugular Tubercle. Turkish Neurosurgery 21: 181-190.

Bass W (1995) Human Osteology. A Laboratory and Field Manual. Missouri Archaeological Society, Columbia.

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 Anatomica 130:344–347.

Franklin D, Cardini A, Flavel A, Kuliukas A (2013) Estimation of sex from cranial measurements in a Western Australian population. Forensic Science International 229:158.e1-158.e8.

Franklin D, Freedman L, Milne N (2005) Sexual dimorphism and discriminant function sexing in indigenous South African crania. Homo 55:213–228.

Gapert R, Black S, Last J (2009a) Sex determination from the foramen magnum: discriminant function analysis in an eighteenth and nineteenth century British sample. International Journal of Legal Medicine 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. American Journal of Physical Anthropology 138:384-94.

Gapert R, Black S, Last J (2013) Test of age-related variation in the craniometry of the adulthuman foramen magnum region: implications for sex determination methods. Forensic Science, Medicine and Pathology 9:478-488.

Gapert R, Last J (2005) The adult human cranial bone: measurement variance and observer error. In: Zakrzewski SR and Clegg M, editors. Proceedings of the Fifth Annual Conference of the British Association for Biological Anthropology and Osteoarchaeology. BAR International Series. Archaeopress, Oxford, pp 119-22.

Guidotti A (1984) Morphometrical considerations on occipital condyles. Anthropologischer Anzeiger 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. American Journal of Physical Anthropology 70:203–208.

Ion A (2011) A brief overview of “Francisc J. Rainer” Human Osteological Collection. Annuals of Romanian Anthropology 48: 24–32.

Kajanoja P (1969) Sex determination of Finnish crania by discriminant function analysis. Archives of Oral Biology 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 Jr PJ (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, Çiçekcibaşi 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, Tetik S, Arda MN (2005) Morphometric analysis of human occipital condyle. Clinical Neurology and Neurosurgery 107: 191-199.

Perini TA, Oliveria G, Ornellas J, Oliveria F (2005) Technical error of measurement in anthropometry. Revista Brasileira de Medicina do 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. International Journal of Legal Medicine 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.

Suazo GIC, Russo PP, Zavando MDA, Smith RL (2009) Sexual dimorphism in the foramen magnum dimensions. Internal Journal of Morphology 27: 21-23.

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.

Uthman AT, Al-Rawi NH, Al-Timini JF (2012) Evaluation of foramen magnum in gender determination using helical CT scanning. Dentomaxillofacial Radiology 41:197-202.

Wescott D, Moore-Jansen P (2001) Metric variation in the human occipital bone: forensic anthropological applications. Journal of Forensic Science 5: 1159–1163.