The Reliability of an Externally-Derived Method for Sexing Skulls When Applied on Archaeological Samples from Banton Island, Philippines

Myra Grace Lara

Introduction

The correct determination of the sex of a skeleton has implications on proper archaeological interpretation such as in burial practices, ancient mortality and demography, social organization and structure. It has become standard procedure in every skeletal analysis to determine sex in addition to estimations of age-at-death, physique, and identification of pathology and trauma. Among the portions of the human skeleton, however, the skull is the one that has received the most attention, the one that is usually retrieved and preserved, especially in earlier archaeological excavations, and the most recognizable skeletal form even to the non-specialist.

Sex determination of a skeleton based on the skull is only considered secondary to that based on the pelvis in terms of accuracy and reliability. For example, Meindl *et al.* (1985), using a set of modern skeletons of known sex, found that 96.0 % of the material could be correctly sexed if the pelvis is used alone but only 92.0 % of this if the skull is instead used. Yet most of the time using the skull is the only option available when the pelvis is fragmented or absent and/or only the skull is retrieved and intact. The most widely applied methods in sexing the skull rely on morphological features, although metric methods have also been developed the method had been derived from (e.g. Stewart 1979; St Hoyme & Iscan 1989; Bass 1995). This premise is based on variation, usually unknown, existing among populational groups separated in time and space. Whether to apply a certain method or not to a populational group under study then becomes a predicament at both ends: Applying a method to a populational group to which it is not actually applicable would have serious implications in the interpretation of results, but failing to apply a method that essentially is applicable is to lose potentially important information.

The present study is an attempt to address the problem presented above. Specifically, it is an attempt to determine whether a particular externally generated sexing method used for skulls, which is the one presented in A&N, could be applied to a Philippine sample known to be from a single group. For this purpose, the cranial series from Banton Island, Romblon, Philippines will be used here.

Materials

Among the skeletal materials available in either the Philippine National Museum or Archaeological Studies Program University of the Philippines, the skeletal series from Banton Island represents one of the most numerous and, for this reason, yielded a number of studies (e.g., Pietrusewsky 1983; Fuentes & Malan 1989; Suzuki *et al*. 1993; Medrana 2005). Although many of the crania in this series exhibit artificial cranial deformation, many are also normal (that is, were not artificially deformed) while some others, although expressing some form of asymmetry in the cranial vault, could not be easily resolved whether they had been artificially deformed or not (see Table 1). The cranial series are supposed to be part of the same material found by Evangelista (1966) in the caves of Guyangan along with other artifacts, including a number of wooden coffins. Some of the skeletal materials were noted as associated with four of these wooden coffins (Archaeological Field Specimen Inventory Record 1961, 1966, n.d.). Medrana (2005) placed the assemblage of human remains and associated artifacts between the dates 960 CE and 1644 CE based on the ceramics recovered with them.

There were originally 28 crania found from three storage rooms in the Philippine National Museum. Many crania were noted to have some sort of labels written on their frontal bones, their left or right temporals, occipital, or left or right maxilla (see Table 1), which suggest that they were marked at different times by different researchers. These different labels were used here to determine whether a particular cranium is from Banton or not and retained and used by the present researcher. There were probably more crania from the Banton series which were not found in the storage rooms and, therefore, not included in the present series.

A study made by Fuentes & Malan (1989) of crania from different sites listed at least five which were not accounted for in the present series: these are 22, 30, GN-II-4, 19 and GN-I-3.

Two of the 28 crania were obviously sub-adults, specifically children, probably with ages-at-death between 2-6 years, and, thus, were excluded from this study. Of the 26 crania originally examined, eight exhibit some degree of artificial deformation, which were also excluded, bringing down the sample to 18. Unfortunately, the statistical method used here could not handle missing data, and two of the 18 remaining samples were lacking the occipital portions to be scored in this region. Hence, these were also excluded, finally bringing down the sample to 16 crania.

Of the remaining 16 crania from the Banton series examined, two more were found to be sub-adults, this time juveniles with estimated ages-at-death ranging between 12 and 18 years, based on a combination of the non-eruption of the third molars and the non-union of the basilar suture. Practices in past studies relating to demography which have used 15 years as the beginning of the adult years, this being the average age of menarche in many cultures (e.g., Angel *et al.* 1986; Pietrusewsky & Douglas 2002), may allow for their inclusion in the present series. On the other hand, there were also others who advised on only using remains estimated to be at least 18 years of age as most of the secondary sexual characteristics appear during pubertal period (Beach 1978; Rice 1984), so that very young males retain more feminine or gracile features (Walker 1995). In Table 1, those that are estimated to be between 12 and 18 years were still included but their entries were highlighted for ease of exclusion during analysis if warranted. Their inclusion in the tables is meant to see if effects of young age in the expression of sex characteristics are also observed in the material.

It is a different case with old adults, especially with old females. In old female adults, the hormones affecting the expression of female secondary characteristics decrease at menopausal stage, allowing them to develop more masculine or robust cranial morphology (Walker 1995). The problem now here is in separating the old adults from the young and middle adults. Estimating age through cranial sutures has received suspicion and sometimes deemed inaccurate (e.g., Brooks 1955). Dental wear could have helped but teeth in most of the skulls in the Banton series have been lost postmortem. However, there are three crania whose maxillary dental arches are completely edentulous (Crania 13 has one remaining third molar) and might suggest that they could have belonged to elderly adults. But this assessment rests on weak grounds (old age is not the only cause of loss of all teeth), especially in the absence of other age-markers. For this reason, the assessment of these crania here as old adults is only, at best, provisional. Their entries in Table 1 are also highlighted for the same reason given for very young individuals.

Methods

The method for sex determination from the skull used here has appeared in similar, if not verbatim, forms in a number of protocols and other guidelines for the examination of human skeletons (e.g. Buikstra & Ubelaker 1994; Moore-Jansen *et al.* 1994; Steadman 1999; White & Folkens 2000; Crane & Carpenter 2006; Steckel *et al.* 2006). But, probably unknown to many, the method was actually an adapted version from a method presented in A&N. The method presented there, in turn, was developed from a number of different workers (e.g. from Broca 1879). The method originally presented in A&N involved a number of regions not only in the skull but also in the pelvis and femur. But in the skull, among the regions considered are the following: (1) glabella and superciliary arches, (2) mastoid processes, (3) external occipital protuberance, (4) supraorbital margins and orbital aperture, (5) mental eminence, (6) the frontal and parietal eminences, (7) occipital squama, (8) zygomatic arch, (9) malar surface, (10) body of the mandible, (11) mandibular angle and (12) mandibular condyles. Of these regions, the first five are considered the most important in determining sex from the skull and are given double the weights ascribed to the rest of the regions. Acsadi & Nemeskeri (1970) used the following scores to describe the different degrees (in parentheses) of sex expression in those regions: +2 ('hypermasculine'), +1 ('masculine'), 0 ('indifferent'), –1 ('feminine') and –2 ('hyperfeminine').

In the method that would later develop from A&N, the same five morphological regions as mentioned above are used but with slight modifications: here the shape of the orbital aperture is no longer considered and, in place of the scores '–2' to '+2', the scores '1' to '5' are used instead. The descriptions for these scores remained the same (see below), however, and essentially the same procedure of assigning scores is also followed. In the revised method, each region is assigned a score from '1' to '5', the minimal and maximal scores respectively, meant to correspond to the 'weakest' and 'strongest' forms of each morphological region. Descriptions of these extreme forms, to which the minimal and maximal scores should be assigned, are provided for each morphological region while intermediate forms, corresponding to scores '2', '3' and '4', are illustrated (see Appendix 1). Compared to the descriptions of scores given in A&N, the scores here correspond to the following: a score of '1' indicates a female, a '2' a probable female, a '4' a probable male and a score of '5' a male. A score of '3' denotes an 'indeterminate' judgment, when the degree of expression of a particular cranial trait was not sufficiently gracile to be judged as female, nor sufficiently robust to be judged as male. For brevity and for lack of a formal name to ascribe to this method, this will be referred here as the revised Acsadi & Nemeskeri, or A&N, method and the method presented in A&N as, simply, the A&N method.

The A&N method was mainly intended for application to European populations although its applicability to non-European samples is not discounted. In fact, Acsadi & Nemeskeri (1970) tested the method on an Afro-American populational group in addition to the 13 archaeological skeletal series from USSR, Hungary and Austria. Presently, this method, or its revised version, has also been widely used in many skeletal analyses in forensic and archaeological contexts (e.g. Steadman 1999; Graw 2001).

The main goal of this research is to determine if the above method, developed from a European population, is reliable when applied to the Banton cranial series. The general plan in achieving this is as follows:

First, the method is applied to the Banton samples by assigning scores to the morphological regions according to the descriptions in the revised A&N method. For this purpose, the one presented in Buikstra & Ubelaker (1994) is used. The specific procedures followed in the assignment of scores are further described below.

Then, the range of variation in the expression of each morphological region, as described in the method, is compared to that observed in the Banton samples. The approach of determining whether similar range of variation exists in the Banton sample will be simple: this will only involve finding out if the 'weakest' and 'strongest' expressions of each morphological region found in the Banton samples conform to the descriptions of the minimal and maximal expressions of the same morphological regions according to the revised A&N method.

The reliability of the method is lastly evaluated by determining if its application to the Banton series had produced consistent results, or if the scores assigned to the morphological regions are consistent with each other. In order to examine consistency among scores, two methods are followed here. The first method is direct and simple, and describes consistency, also called 'agreement', among scores in terms of the absolute distances among them and the proportion of the times two scores in the same cranium agree to the times they disagree. A definition of which scores agree and which do not are given in the next section. The second method is in terms of using a model to quantify consistency or agreement. For this purpose, intraclass correlation, a type of inter-rater reliability test, is used.

Specific procedures in assigning scores

Of the five cranial morphological regions considered in the revised A&N method, only four could be employed here. The reason is obvious: mental eminence is a site in the mandibular corpus and, although there are a few mandibles found with the crania in the National Museum storage rooms, none of them is clearly associated with any of the crania.

The remaining 16 crania were scored according to the four remaining cranial morphological regions of nuchal crest, mastoid process, supraorbital margin and glabella. Scoring of each morphological region was done through seriation. Seriation involves the simple arrangement of crania, where a cranium is placed in one of 16 positions, from the least expression of a particular morphological region to a massive expression. The judgment of what is 'least' and 'massive' expressions, and what is in between, had been guided by the descriptions given for the method in one reference (Buikstra & Ubelaker 1994). Hence, for instance, in seriating the crania according to the surface rugosity of the nuchal crest, a 'smooth with no bony projections' defines a minimal expression, while one that 'forms a well-defined bony ledge or hook' is considered very massive. Scores then are assigned to the crania by dividing them into discrete groups that conform to the five scores. Seriation and scoring of crania are done separately for each morphological region.

In practice, the usual approach in assigning scores is by direct application, whereby crania are scored one at a time. But assigning of scores through seriation minimizes intra-observer error (Meindl *et al.* 1983; Lovejoy *et al.* 1985; Meindl & Lovejoy 1989, as they used seriation in scoring pubic symphysis) and provides an opportunity to resolve problematic cases while all the bones are still laid out and available for comparisons.

Table 1. Scores assigned to morphological regions in the Banton cranial series (see Attachment 1 for details of the method):

Crania	Other labels	Description	Mastoid	Glabella	Orbit	Occipital
F	"GN III I"	No obvious deformation but with asymmetry	2	2	1	1
36*	"GN III 2"	No obvious deformation	3	2	2	1
6	"GN-I" "Banton Is GN-I"	No obvious deformation	3	2	2	2
8	"Banton Is GN-I"	No obvious deformation	3	2	2	2

Crania	Other labels	Description	Mastoid	Glabella	Orbit	Occipital
35**	"Banton GN-I"	Not deformed	2	2	3	2
37*	"Banton" "GN III 3"	No obvious deformation	3	3	1	3
34	"Banton GN-I"	No obvious deformation but with asymmetry	3	4	1	2
2	"Banton I GN-I" (with weathering)	Not deformed	2	4	2	3
5	"Banton Is GN-" (with erasure?)	No obvious deformation	5	3	2	3
33	"Banton GN-I"	No obvious deformation but with slight asymmetry	3	3	3	4
45**	-	No obvious deformation	4	4	5	2
Guyangan 1	"Guyangan I Banton Is"	No obvious deformation but with asymmetry	3	5	4	5
Guyangan III-13	"GN III 13"	Not deformed	3	5	4	5
9	"Banton Is. GN-1"	No obvious deformation but with asymmetry	5	5	3	5
31	"GN III III"	No obvious deformation	5	5	5	4
13**	"Banton Is GN I"	No obvious deformation	5	5	4	5
29	"Banton"	Not deformed	2	1	2	-
11	"Banton Is. GN-I"	Not deformed	3	3	2	-

Notes: Specimens (except in the last two rows) are arranged according to the average of their scores. The entries for the juvenile specimens are marked with an *, and those for old adults are marked with **;

Interpretation of scores: **1** – female, **2** – probable female, **3** – unknown, **4** – probable male, **5** – male.

Results and Discussion

The scores assigned to each morphological region are given in Table 1. Photographs of crania representing each score in each morphological region are shown in Appendix 2.

Note that Cranium 36, one of the two juveniles, is consistently scored as either a '1' or a '2' in at least three cranial morphological regions making it possible that Cranium 36 is a young male that retain female features at time of death. Despite this possibility, the specimen is not removed from the series since its exclusion does not significantly change the results. With regard to crania estimated to be belonging to old adults, only one had been consistently evaluated as a male or a probable male in at least three morphological regions (Cranium 13) while one had been consistently evaluated as a probable female in at least three morphological regions (Cranium 35). The last old adult (Cranium 45), is a different case having been evaluated as a male or a probable male in three morphological regions but a probable female in one (see below).

Minima and maxima expressions in the Banton series and in the revised Acsadi & Nemeskeri (A&N) method

The minimal expression of supraorbital margin and nuchal crest, as described in the revised A&N method, is found in four crania (F, 36, 37 and 34). For example, Crania F, 37 and 34 exhibit supra-orbital margins that are very sharp, similar to an "edge of a slightly dulled knife". Also, Crania F and 36, exhibit nuchal crests that are smooth and without bony projections. On the other hand, the least expressions of glabella and mastoid processes are not observed in any of the samples (Figure 1).

Nonetheless, the maximal expression of all four morphological regions were noted in many crania. For example, the nuchal crest in Crania 9 and Guyangan III-13 are massive and form a well-defined hook, consistent with the description of Score '5' in the revised A&N method. In addition, the supra-orbital margin of Crania 45 and 31 are thick and rounded and almost "pencil-like" in its curvature, as also described in the revised A&N method. Very pronounced glabellar region and very large mastoid processes of the magnitude stated in the revised A&N method are also found exhibited in Crania 9 and 31.

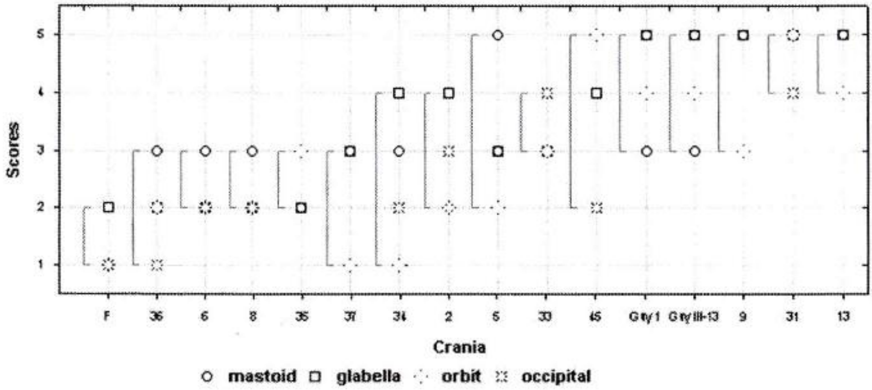


Figure 1. Plots of scores for each morphoogical region in each cranium

Absolute distances among scores

Absolute distances between pairs of scores are given in Figure 2. Since there are six possible combinations of four morphological regions and there are 16 crania considered here, 96 distances (from 96 pairs of scores) are generated here.

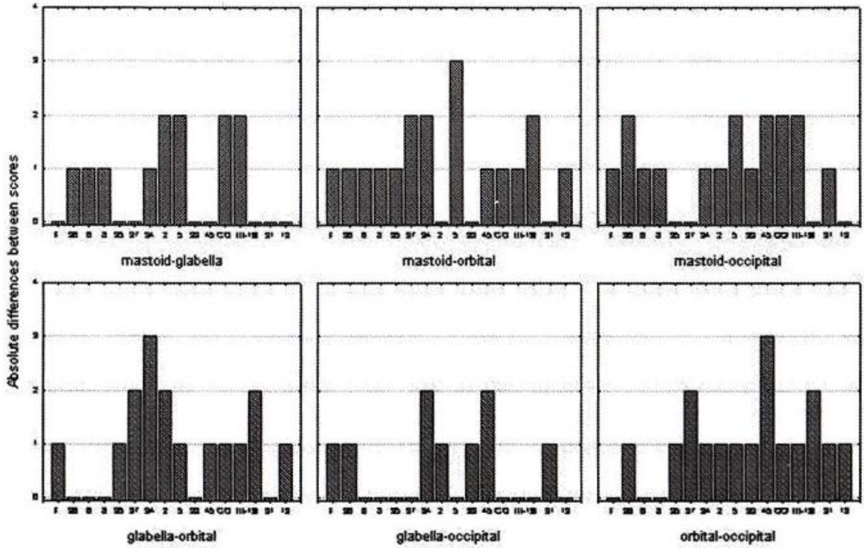


Figure 2. Plots of differences among scores assigned to morphological regions

The following are taken as guides in determining the consistency of score assignments across each cranium. Their formulation is based on the interpretation of each score as intended in the revised A&N method (i.e. '1' is female, '5' is male):

(1) Equal scores, or scores with a distance of 0, are consistent. This means that a score of '1' in one region is supported by a score of '1' in another region;

(2) Adjacent scores, or scores with a distance of 1, are also consistent. Hence, a score of '1' in one region supports (and is supported by) a score of '2' in another region. Similar principle applies to scores of '4' and '5'.

(3) Scores with a distance of 2 are not consistent (similarly in Graw 2001) except when one of the scores is '3'. This means that a score of '1' or '5' in one region is not inconsistent with a score of '3' in another region. In fact, a score of '1', '2', '4' or '5' combined with a score of '3' is not inconsistent. The reason for this is that an indeterminate score, while not supporting a result of any other scores, is not contradicting it. A score of '2' in one region is, however, inconsistent with a score of '4' in another region.

(4) Scores with a distance of 3 or 4 are not consistent. Possible combinations are '1' and '4', '2' and '5' and '1' and '5', which is the most extreme case of inconsistency.

Of the 96 pairs of scores, 32 pairs have a distance of 0 (which means the pairs are equal scores), 42 pairs have a distance of 1 (which means the pairs are adjacent scores), 19 pairs have a distance of 2 and three pairs have a distance of 3 in their scores. It is important to note also that no pair produced a distance of 4. Five of the 19 pairs with a distance of 2 are inconsistencies, occurring between a score of '2' and '4'. Hence, of the 96 pairs of scores, only eight, or 8.33%, are inconsistencies.

The inconsistencies occurred in four crania: 34, 2, 5 and 45. Note that three of these crania (34, 5 and 45) are the ones with the widest range of scores (extending to four scores, Figure 1). Four of these inconsistencies occurred in two crania (Crania 34 and 2 have each two inconsistent pairs of scores) and are caused, in both, by a relatively robust glabella compared to the other regions. Three inconsistent pairs of scores occurred in only one cranium, Cranium 45, and are caused by a relatively gracile occipital compared to the rest of the regions. Cranium 45 is actually one of the crania that is provisionally assessed here as an old adult (if basis used for ageing is to be relied upon, see Table 1), but whether this has caused the differing expressions of its regions is not known, especially when it is also possible that the differing expressions could simply be part of normal variation.

Intraclass correlation

Reliability, as a term in statistics, is defined as the correlation of a 'rater' with a hypothetical one that truly measures what it is intended to measure. A rater can be an instrument, a questionnaire or, in the present case, a method. Since the true 'rater' could not be obtained, reliability is estimated in different ways, among them is through an assessment of the agreement of results produced by the rater. This is called *inter-rater agreement*. An inter-rater agreement test is based on the correlation of scores given by two or more raters for the same set of objects. A correlation coefficient typically used to measure inter-rater agreement is the *intraclass correlation coefficient* (ICC) (Shrout & Fleiss 1979). ICC may be understood as the ratio of between-group variance to total variance. In simplified terms, a low variation among scores assigned to subjects, produced by the assignment of the same score by the raters, will approach a correlation coefficient of 1.00. Thus, unlike the non-parametric correlation coefficients Spearman rho or Kendall tau, or Kendall's coefficient of concordance, which only evaluates agreement in ranks, intraclass correlation takes into consideration both the ranks and differences among the scores (Wuensch 2003).

The goal in this study is to assess the agreement in the scores assigned to the different morphological regions across all crania using intraclass correlation with morphological regions treated as 'raters'. The derivation of intraclass correlation coefficients uses the two-way mixed model wherein the raters (morphological regions) are fixed (not a random sample of all possible raters) but the objects to be rated (crania) are a random sample. This, in turn, means that the ICC could not be generalized beyond the raters considered here. Both types of ICC computation are also used: one in which the absolute agreement in scores (where scores are identical) is measured, and another in which only consistency in scores (where scores are relatively similar) is measured. This is computed for combinations of two, three and four morphological regions. Since there are four morphological regions considered in this study, what are produced are six, four and one combination of morphological regions, respectively. The intraclass correlation coefficients for combinations of two and three morphological regions are given in Tables 2 and 3. The coefficients at the top row of each cell are those generated by using consistent agreement in scores, those at the bottom had used absolute agreement.

Table 2: Intreclass Correlation Coefficients of Mastoid, Glabella and Orbit to Glabella, Orbit and Occipital

	mastoid	Glabella	Orbit
glabella	0.690 (p < .015) 0.701 (p < .015)		
orbit	0.639 (p < .029) 0.599 (p < .029)	0.771 (p < .004) 0.709 (p < .004)	
occipital	0.646 (p < .026) 0.648 (p < .026)	0.901 (p < .000) 0.881 (p < .000)	0.719 (p < .010) 0.720 (p < .010)

Table 3: Interclass Correlation Coefficients of Orbit, Orbit/Occipital, Occipital to Mastoid, Mastoid/Glabella and Glabella

	orbit	orbit/ occipital	occipital
mastoid		0.755 (p < .000) 0.745 (p < .000)	
mastoid/ glabella	0.783 (p < .000) 0.755 (p < .000)		0.831 (p < .000) 0.828 (p < .000)
glabella		0.858 (p < .000) 0.838 (p < .000)	

Intraclass correlation coefficient is interpreted in the same way as the Cohen's Kappa coefficient (Fleiss & Cohen 1973). By convention, a Cohen's Kappa that is within 0.40 and 0.59 suggests moderate inter-rater reliability; one that is within 0.60 and 0.79 is substantial while a Kappa coefficient that is equal or larger than 0.80 is outstanding (Landis & Koch 1977).

The values of coefficients generated for combinations of two morphological regions indicate outstanding agreement between glabella and occipital and substantial agreement between the rest of the pairs of morphological regions, with all results significant at the $p < 0.03$ level. This is the case whether absolute agreement or only relative agreement among scores is considered in the computation. Although the coefficients generated when absolute agreements are taken into account are lower than those generated when only consistency is considered, the two sets of coefficients are not so much different. This indicates that the number of consistent scores is close to the number of tied scores. In addition, the coefficients follow the same, general order of strength in both types of ICC computation. Using consistency of scores in the computation, the high agreement between glabella/occipital is followed by glabella/orbit, orbit/occipital, glabella/mastoid, mastoid/occipital and mastoid/orbit. With absolute agreement between scores taken into account, however, the occipital/orbit pair has a higher correlation than the glabella/orbit, indicating that there are more tied scores between orbit and occipital than orbit has with glabella. That mastoid occurs in pairs with the lower coefficients might suggests that, among the four regions, mastoid could be the least correlated to the other three. In the same way, the occurrence of glabella in pairs with the higher coefficients suggests that the glabella is the most correlated among the four regions.

The correlations among the morphological regions become stronger when combinations of three regions are considered, suggesting that consistency increases when more 'raters' are taken into account. All strengths here are also significant, but at a higher level of $p < 0.00$. As in the observation made when only pairs of regions were considered, the correlations are weaker whenever the mastoid is included but stronger when the glabella is among the regions.

When all four morphological regions are considered altogether, the value generated for the intraclass correlation coefficient becomes much stronger. The ICC coefficient is 0.851 when only consistency among scores is considered and 0.838 when absolute agreements among scores are accounted for, both of which is significant at the $p < .000$ level. This implies outstanding agreement among scores when all morphological regions are taken simultaneously.

Conclusion

Except for the minimal expression in glabella and mastoid, the minimal and maximal expressions of all morphologies, as described in the revised A&N method, are found to be exhibited in the Banton sample. The absence of score '1' in glabella might have been produced by sampling. Cranium 29, which was excluded in the sample (since it lacked a score in one region) but belonging to the Banton series, exhibit a glabella that is flat and smooth. The non-occurrence of a minimal expression of mastoid could also be due to sampling. At least three crania that also belong to the Banton series are noted to exhibit very small mastoid processes. The crania are, however, artificially deformed and whether the gracile expression of this morphological region could have been affected by this practice is yet to be investigated.

The strong consistency of scores in all morphological regions across each cranium is an important result. This means that the scores in the four morphological regions are correlated in each cranium, where a score in one region implies a similar score in another. Similar scores, by definition in the revised A&N method, imply similar sex attribution in the four morphological regions. If similar sex is being attributed in the four morphological regions, then their use should allow for the separation of males and females in the Banton series.

However, considering that the use of the four morphological regions does separate males and females in the present series, there might be questions whether this separation would be accurate. Tests of accuracy usually require testing the results against the true sex of the samples, which, unfortunately, are unknown for all Banton crania. One obvious reason is that the materials were derived from archaeological contexts and antemortem information regarding the sexes of the individuals is, therefore, not available. Another reason is that none of the crania have retained their association to postcranial skeleton, which means that an assessment of their sex, when made here, could not be possibly verified with those, for example, evaluated from the pelves.

Yet, the result of consistency here could only imply that either the four morphological regions are all *consistently* leading to correct results or *consistently* leading to incorrect findings. Any method that consistently leads to incorrect results would most probably do so by chance. The amount of consistency here is achieved at too high a statistical significance level to have occurred by chance alone. It might be then just a matter of extending the conclusion that if the results produced by the four morphological regions have been consistent, and have been that way not by chance, then they could be accurate as well.

Finding that the range of variation defined in the method is found in another populational group suggests that sexual dimorphism may exist in the same degree in, at least, some populational groups more often than one might be willing to consider. But the reliability of an externally derived sexing method in the Banton cranial series has a practical implication also. It raises the question whether there is a need to create standards for every populational group that comes under study when methods already in place could be used.

The main contribution of the A&N method, or its revised version, lies in its emphasis that sex characteristics in the skeleton manifest in a gradation of forms and not on clearly defined, or discrete, states. Sex differences will always exist in any populational group, and a gradation of differing forms, with two opposing extremes, will always exist in any skeletal series that one encounters. What the A&N method offers is guidance regarding identification of this variation, as in which in the series is to be identified as the 'most feminine' and which is the 'most masculine' form (In one protocol (Crane & Carpenter 2006), forms more extreme than '1' and '5' are even considered, which should be scored as '1' and '5', respectively). It is this identification of extreme forms that are most useful in sex attribution, especially when dealing with isolated crania where variation of forms is not readily observable. The extreme forms that A&N describes have been observed in archaeological and recent samples from different geographic regions. So far, in addition to the Banton cranial series, the method had been tested on recent U.S. skulls of African descent (Acsadi & Nemeskeri 1970) and recent (forensic) skulls from southwestern Germany (Graw 2001). Thus, although an externally generated method, the A&N method can be reliably used for local populations. A reliable method already in place, attempts to establish separate standards for local populations might become redundant.

There are many other methods used in skeletal analysis that were generated from specific populations, not only for sex determination but for age estimation as well (e.g. the Suchey-Brooks method (Brooks & Suchey 1990) used for ageing adult material based on the pubic symphysis and derived from U.S. samples). Determining whether, at least, some of these have wider application outside the base populations will reap the same benefits of not only saving resources that are otherwise spent in establishing standards but also help reveal the degree of inter-populational variation that exists for the different biological attributes of, among others, sexual dimorphism and rate and pattern of ageing.

References

Acsadi, G., and I. Nemeskeri. 1970. *History of Human Life Span and Mortality*. Budapest: Akademiai Kiado.

Angel, J. L., J. M. Suchey, M. Y. Iscan, and M. R. Zimmerman. 1986. "Age at death from the skeleton and viscera," in *Dating and age determination in biological materials*. Edited by M. R. Zimmerman and J. L. Angel, pp. 179-220. London: Croom Helm.

Archaeological Field Specimen Inventory Record. 1961. Guyangan I. Manuscript of the National Museum of the Philippines, Manila.

Archaeological Field Specimen Inventory Record. 1966. Guyangan Tomb (GN-III). Manuscript of the National Museum of the Philippines, Manila.

Archaeological Field Specimen Inventory Record. n.d.. Guyangan II. Manuscript of the National Museum of the Philippines, Manila.

Bass, W. M. 1995. *Human Osteology: A Laboratory and Field Manual*, fourth edition. Columbia: Missouri Archaeological Society.

Beach, F. A. 1978. "Human sexuality and evolution," in *Human Evolution: Biosocial Perspectives*. Edited by S. L. Washburn and E. R. McCown, pp. 123-53. Menlo Park, California: Cummings.

Broca, P. 1879. "Instruction craniologique," in *Dictionaire encyclopédique des sciences médicales*. Edited by D. Dechambre, pp. 642. Paris: Asselin/Masson.

Brooks, S. T. 1955. Skeletal Age at Death: the Reliability of Cranial and Pubic Age Indicators. *American Journal of Physical Anthropology* 13:567-597.

Brooks, S., and J. M. Suchey. 1990. Skeletal age determination based on the os pubis: a comparison of the Acsadi-Nemeskeri and Suchey-Brooks method. *Human Evolution* 5:227-38.

Brothwell, D. R. 1965. *Digging up Bones: the excavation, treatment and study of human skeletal remains*. London: British Museum of Natural History.

Buikstra, J., and D. H. Ubelaker. 1994. *Standards for Data Collection from Human Skeletal Remains Proceedings of a Seminar at the Field Museum of Natural History*. Fayetteville: Arkansas Archaeological Survey Research Series, No. 44

Crane, S., and R. Carpenter. 2006. *Poultron Research Project Skeleton Manual*. Liverpool.

Evangelista, A. E. 1966. Report on an archaeological expedition to the island of Banton, Romblon province, Philippines. National Museum of the Philippines.

Fleiss, J. L., and J. Cohen. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement* 33:613-619.

Fuentes, J. C., and I. Malan. 1989. Craniometric Measurements and Analysis of Some Skulls from Philippine Archeological Sites. National Museum of the Philippines.

Giles, E. 1964. Sex determination by discriminant function analysis of the mandible. *American Journal of Physical Anthropology* 22:129-36.

Giles, E., and O. Elliot. 1963. Sex determination by discriminant function analysis of crania. *American Journal of Physical Anthropology* 21:53-68.

Graw, M. 2001. Significance of the Classical Morphological Criteria for Identifying

Gender Using Recent Skulls. Paper presented at the 9th Biennial Meeting of the International Association for Craniofacial Identification. FBI. Washington, DC. July 24, 2000. *Forensic Science Communications*, 3 (1).

Graw, M., A. Czarnetzki, and H. T. Haffner. 1999. The form of the supraorbital margin as a criterion in identification of sex from the skull: Investigations based on modern human skulls. *American Journal of Physical Anthropology* 108:91-96.

Hanihara, K. 1959. Sex diagnosis of Japanese skulls and scapulae by means of discriminant function. *J. Anthropol. Soc. Nippon* 67:191-97.

Holland, T. D. 1986. Sex determination of fragmentary crania by analysis of the cranial bases. *American Journal of Physical Anthropology* 70:203-208.

Hoshi, H. 1962. Sex difference in the shape of the mastoid process in norma occipitalis and its importance to the sex determination of the human skull. *Okafimas Folia Anat. Jpn.* 38:309-13.

Kajanoja, P. 1966. Sex determination of Finnish crania by discriminant function analysis. *American Journal of Physical Anthropology* 24:29-34.

Keen, H. A. 1950. A study of sex differences in male and female skulls. *American Journal of Physical Anthropology* 8:65-79.

Krogman, W. M. 1962. *The Human Skeleton in Forensic Medicine*. Illinois: Thomas, Springfield.

Landis, J. R., and G. G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33:159-174.

Lovejoy, C. O., R. S. Meindl, R. P. Mensforth, and T. J. Barton. 1985. Multifactorial determination of skeletal age at death: a method and blind tests of its accuracy. *American Journal of Physical Anthropology* 68:1-14.

Martin, E. S. 1936. A study of an Egyptian series of mandibles, with special reference to mathematical methods of sexing. *Biometrika* 28:149-78.

Medrana, J.G. 2005. The People of Ancient Banton: Assemblage composition, population features, and health through osteological analysis. MSc dissertation, University of the Philippines.

Meindl, R. S., and C. O. Lovejoy. 1989. "Age changes in the pelvis: Implications for paleodemography," in *Age Markers in the Human Skeleton*. Edited by M. Y. Ýþcan, pp. 137-168. Springfield, Illinois: Charles Thomas Publisher.

Meindl, R. S., C. O. Lovejoy, and R. P. Mensforth. 1983. Skeletal age at death: accuracy of determination and implications for human demography. *Human Biology* 55:73-87.

Meindl, R. S., C. O. Lovejoy, R. P. Mensforth, and R. A. Walker. 1985. A revised method of age determination using the os pubis, with a review and tests of accuracy of other current methods of pubic symphyseal aging. *American Journal of Physical Anthropology* 68:29-45.

Moore-Jansen, P. M., S. D. Ousley, and R. L. Jantz. 1994. *Data Collection Procedures for ForensicSkeletal Material*: Forensic Anthropology Center, University of Tennessee, Knoxville.

Novotny, V., M. Y. Ýþcan, and S. R. Loth. 1993. "Morphologic and osteometric assessment of age, sex, and race from the skull," in *Forensic Analysis of the Skull: Craniofacial Analysis, Reconstruction and Identification*. Edited by M. Y. Ýþcan and R. P. Helmer, pp. 71-88. New York: John Wiley.

Pietrusewsky, M. 1983. (Computer print-out of cranial measurements and non-metrical observations attached to letter written to Dr. Alfredo E. Evangelista and dated March 8, 1983). National Museum, Manila, Philippines.

Pietrusewsky, M., and M. T. Douglas. 2002. *Ban Chiang, a Prehistoric Village Site in Northeast Thailand: The Human Skeletal Remains (Thai Archaeology Monograph Series, 1)*: University of Pennsylvania.

Rice, R. W. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution* 38:735-42.

Shrout, P. E., and J. L. Fleiss. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86:420-428.

St Hoyme, L. E., and M. Y. Iscan. 1989. "Determination of sex and race: Accuracy and assumptions," in *Reconstruction of Life from the Skeleton*. Edited by M. Y. Iscan and K. A. R. Kennedy, pp. 53-93. New York: Alan R. Liss Inc.

Steadman, D.W. 1999. *Physicians for Human Rights Cyprus Project Forensic Training Manual*. Cyprus. (unpublished)

Steckel, R. H., C. P. Larsen, P. W. Sciulli, and P. L. Walker. 2006. *The Global History Of Health Project Data Collection Codebook*. Ohio.

Stewart, T. D. 1979. *Essentials of Forensic Anthropology, Especially as Developed in the United States*. Springfield, Illinois: Thomas.

Steyn, M., and M. Y. Iscan. 1998. Sexual dimorphism in the crania and mandibles of South African whites. *Forensic Science International* 98:9-16.

Suzuki, H., Y. Mizoguchi, and E. Conese. 1993. Craniofacial Measurement of Artificially Deformed Skulls from the Philippines. *Antropol. Sci* 101:111-127.

Walker, P. L. 1995. "Problems of preservation and sexism in sexing: Some lessons from historical collections for palaeodemographers," in *Grave Reflections: Portraying the Past through Cemetery Studies*. Edited by S. Saunders and A. Herring, pp. 31-48. Toronto: Canadian Scholar's Press.

White, T. D., and P. A. Folkens. 2000. *Human Osteology*, 2nd edition. San Diego, California: Academic Press.

Wuensch, K.L. 2003. *Inter-Rater Agreement*. http://core.ecu.edu/psyc/wuenschk/docs30/InterRater.doc. Accessed on July 12, 2006.

Acknowledgments

I owe the opportunity to analyze the Banton series to Prof. Angel Bautista, who in 2003, gave me permission to access the materials and advice in the choice of samples. The motivation to write the results, after three years, I owe to Dr. Victor Paz, who, in one casual conversation regarding a case in Cebu, remarked that those who do not write their thoughts, for fear of being wrong, are those who whine in pubs when someone else did write and turned out right. I could still turn

out wrong but, at least, there is no more chance that I would end up in some dusky pub drinking my gripes away.

Abstract

A number of methods have already been developed for analyzing unknown human skeletal material, which includes methods for ageing, sexing and estimating stature. Usually, the development of these methods had relied upon particular base populations, thus, their use to outside populations has been generally presumed unreliable. The present study investigates this presumption, by determining the reliability of a particular method when applied to another population. The Acsadi & Nemeskeri method, a sexing method based on the skull and developed from European samples, is applied here to a Philippine sample known to have come from a single group – the cranial series from the Guyangan Caves in Banton Island, the Philippines. The method involves assigning scores to five morphological regions: the nuchal crest, mastoid process, supraorbital margin, glabella and mental eminence. The reliability of this method is evaluated by determining if its application to the Banton series had produced consistent results, or if the scores assigned to the morphological regions are consistent with each other. Two approaches were followed to examine consistency among scores: the first is direct and simple, and describes consistency among scores in terms of the absolute distances among them. The second approach uses a model to quantify consistency and considers the intraclass correlation, a type of inter-rater reliability test, among scores. The amounts of consistency achieved in both approaches are high, indicating strong consistency among scores. The reliability of the Acsadi & Nemeskeri method to the Banton series is then argued for, suggesting that externally-derived methods could have wider applications than many researchers might have been willing to consider.

Appendix 1. Sex determination from the skull according to the revised Acsadi & Nemeskeri (A&N) method (after Buikstra & Ubelaker, 1994).

For adults, the following sexually dimorphic cranial traits will be scored through reference to the Figures 1-5: nuchal crest, mastoid process, prominence of glabella, supraorbital margin and mental eminence. Score each trait independently, ignoring other features. Specific procedures for each of the traits are listed below. The most extreme forms of each feature are defined here, with intermediate grades illustrated visually in Figures 1-5.

Nuchal Crest: View the lateral profile of the occipital and compare it with the diagrams. Feel the surface of the occipital at the midline with your hand and note any surface rugosity, ignoring the con-tour of the underlying bone. Focus upon the rugosity attendant to attachment of nuchal musculature. In the case of minimal expression (score = "1"), the external surface of the occipital is smooth with no bony projections visible when the lateral profile is viewed. Maximal expression (score = "5") defines a massive nuchal crest that projects a considerable distance from the bone and forms a well-defined bony ledge or "hook."

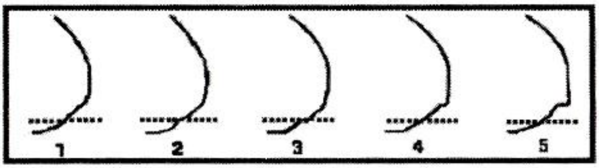


Figure 1: Standard for scoring the nuchal crest.

Mastoid Process: Score this feature by comparing its size with that of surrounding structures such as the external auditory meatus and the zygomatic process of the temporal bone. Mastoid processes vary considerably in their proportions. The most important variable to consider in scoring this trait is the volume of the mastoid process, not its length. Minimal expression (score = "1") is a very small mastoid process that projects only a small distance below the inferior margins of the external auditory meatus and the digastric groove. A massive mastoid process with lengths and widths several times that of the external auditory meatus should be scored as "5."

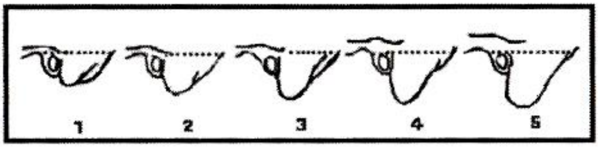


Figure 2: Standard for scoring the mastoid process.

Supraorbital Margin: Begin by holding your finger against the margin of the orbit at the lateral aspect of the supraorbital foramen. Then hold the edge of the orbit between your fingers to determine its thickness. Look at each of the diagrams to determine which it matches most closely. In an example of minimal expression (score = "1"), the border should feel extremely sharp, like the edge of a slightly dulled knife. A thick, rounded margin with a curvature approximating a pencil should be scored as "5."

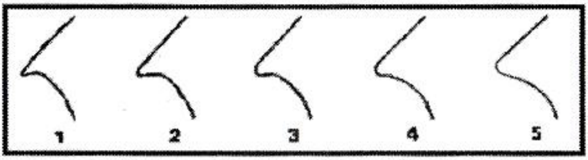


Figure 3: Standard for scoring the supraorbital margin.

Prominence of Glabella: Viewing the cranium from the side, compare the profile of the supranasal region with Figure 4. In a minimal prominence of glabella (score = "1"), the contour of the frontal is smooth, with little or no projection at the midline. Maximal expression involves a massive glabellar prominence, forming a rounded, loaf-shaped projection that is frequently associated with well-developed supraorbital ridges.

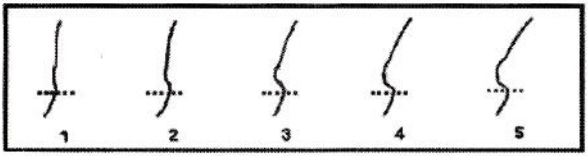


Figure 4: Standard for scoring the prominence of glabella.

Mental Eminence: Hold the mandible between the thumbs and index fingers with thumbs on either side of the mental eminence. Move the thumbs medially until they delimit the lateral borders of the mental eminence. In examples of minimal expression (score = "1"), there is little or no projection of the mental eminence above the surrounding bone. By contrast, a massive mental eminence that occupies most of the anterior portion of the mandible is scored as "5."

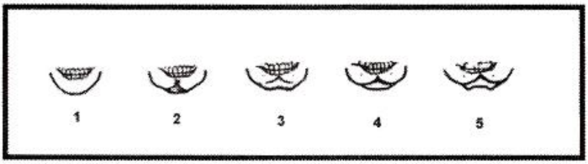


Figure 5: Standard for scoring the mental eminence.

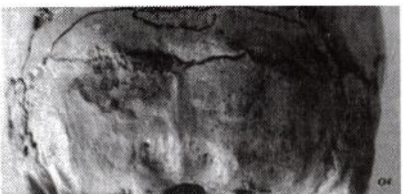
Appendix 2. Representative crania for each score in each morphological region in the Banton series.

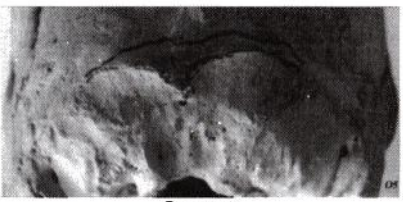
Nuchal crest


Score 1


Score 2

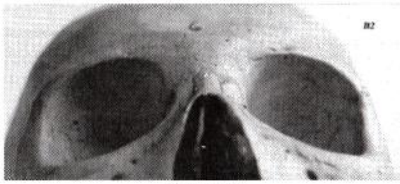

Score 3


Score 4

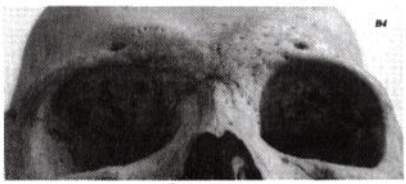

Score 5

Supraorbital margin


Score 1


Score 2

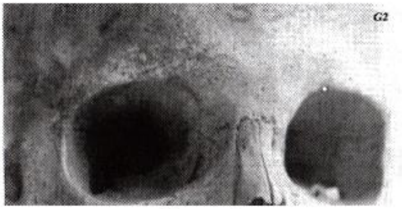

Score 3


Score 4


Score 5

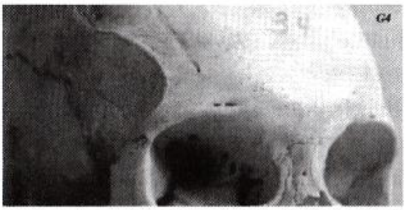
Glabella

Score 1



Score 2

Score 3



Score 4

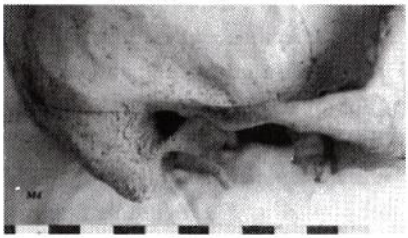
Score 5

Mastoid process

Score 1

No image for Score 2

Score 3



Score 4

Score 5