

# Two arch criteria of the ilium for sex determination of immature skeletal remains: A test of their accuracy and an assessment of intra- and inter-observer error

Hugo F.V. Cardoso<sup>a,b,c,\*</sup>, Shelley R. Saunders<sup>c</sup>

<sup>a</sup>Departamento de Antropologia, Universidade de Coimbra, Rua do Arco da Traição, 3000-056 Coimbra, Portugal

<sup>b</sup>Faculdade de Medicina, Universidade do Porto, Al. Prof. Hernâni Monteiro, 4200-319 Porto, Portugal

<sup>c</sup>Department of Anthropology, McMaster University, Chester New Hall 524, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4L9

Received 1 November 2007; received in revised form 22 January 2008; accepted 29 January 2008

Available online 17 March 2008

## Abstract

Although the assignment of sex to immature skeletal remains is considered problematic, some traits have been considered useful for both forensic and bioarchaeological applications. One such trait is the arch criterion found in subadult ilia, which is defined relative to the greater sciatic notch-auricular surface area. In adults, the composite arch has also been described in relation to this area and has proven relatively successful in sex determination. This study offers an examination of the accuracy of the arch criterion and the composite arch in determining the sex of subadult skeletal remains, and an assessment of intra- and inter-observer scoring error. A sample of 97 skeletons of known sex and age (<15 years) from the Lisbon collection (Portugal) were selected and the traits were scored by three observers on orthogonal photos of each ilium. In general the agreement within (67.7–88.5%) and between (50.5–76.3%) examiners was poor and overall accuracy (26.7–52.6%) did not meet the expectations of that reported in previous studies. The authors suggest that this derives from great variation in morphology, difficulties in interpreting criteria and possibly a lack of association between the expression of the traits and sex. Careful examination of sex-related morphology in the immature skeleton and additional blind tests of so-called useful traits should continue to be carried out.

© 2008 Elsevier Ireland Ltd. All rights reserved.

**Keywords:** Forensic science; Subadults; Sex determination; Ilium; Arch criterion; Composite arch

## 1. Introduction

There has been widespread recognition of the inability to determine sex in immature skeletal remains [1]. Several attempts to develop methods to determine the sex of subadult skeletons have relied on the identification of adult sex-related features. This relates to the expectation that adult-like differences begin to show even in infancy. Since the pelvis is the most sexually dimorphic part of the adult skeleton it would make sense to look for dimorphism in the subadult pelvis and there is a long literature on this subject [2–8]. One area of the skeleton that has received considerable investiga-

tion is the greater sciatic notch-auricular surface area in the pelvis. In adults, the composite arch is defined in relation to this area and has been successfully utilized in sex determination. The composite arch was proposed by Genovés [9], who obtained 80% of correct allocation accuracy for males and 88% for females in a mixed sample of English origin. Similar correct sex allocation accuracies were obtained by Bruzek [10], 67% for males and 92% for females in a French sample ( $N = 162$ ), and 78% for males and 87% for females in a Portuguese sample ( $N = 240$ ). In the subadult skeleton, the arch criterion is also defined in relation to the greater sciatic notch-auricular surface area and has also been relatively successful in sex determination. The arch criterion was proposed by Schutkowski [11] and is very similar to the composite arch criterion. A correct sex allocation accuracy of 81.5% for males and 60% for females was obtained by Schutkowski [11] in a sample of 61 subadult skeletons from birth to 5 years of age, using the Spitalfields collection in

\* Corresponding author at: Departamento de Antropologia, Universidade de Coimbra, Rua do Arco da Traição, 3000-056 Coimbra, Portugal.  
Tel.: +351 239 829 051; fax: +351 239 823 491.

E-mail address: [hcardoso@antrop.uc.pt](mailto:hcardoso@antrop.uc.pt) (H.F.V. Cardoso).

London. Sutter [12] further tested the arch criterion on a sample of 85 pre-Colombian mummies of known sex from Chile and 76.9% correct sex allocation was achieved for males and 85.7% for females, in the same age range as in Schutkowski's [11] sample. When the ages were extended to 15 years of age, Sutter [12] achieved 68% correct sex allocation accuracy for males and 91.9% for females.

This overall consistency in sex-specific traits of the greater sciatic notch-auricular surface area in both adult and subadult remains, suggests good reliability of this area and that the traits proposed are useful for sex determination of skeletal remains from birth to adulthood. However, using the composite arch in adults, Novotný [13] obtained only 67% of correct sex allocation for males and 56% for females in a Czech sample ( $N = 226$ ). Additionally, using the arch criterion in subadults, greater correct sex allocation accuracy was obtained for females in Sutter's [12] test, while it was greater for males in Schutkowski's [11] study. These results suggest that, although some consistency in sex-specific traits has been observed in the greater sciatic notch-auricular surface area, significant variation may exist in adult and subadult pelvic morphology of different populations for these traits to be universally applied with sufficient accuracy. Given that the evidence for population variation in the greater sciatic notch-auricular surface area of subadults has been assessed only to a limited degree, research using additional documented sex subadult collections is important to determine whether traits observed in this area are sufficiently accurate and reliable to be used with other samples. The Lisbon collection is a series of Portuguese documented skeletons with a relatively large subadult segment, and provides the unique opportunity to test the reliability of the greater sciatic notch-auricular surface area in determining the sex of immature human remains. Besides population variability as an explanation for variability, variation in accuracy of the greater sciatic notch-auricular surface area in determining sex may result from the inability of different researchers to score the same trait consistently. In fact, a major source of problems in sex determination is that many workers have difficulties perceiving and scoring the standards for traits established by others, even when pictures are provided [14]. Obviously, these problems raise some concerns as to the accuracy of the so-called useful traits.

The initial goal of this study was to test the accuracy of the above-mentioned two traits (composite arch and arch criterion) for sex determination of subadult ilia, but it became increasingly apparent that the scoring of these features was not sufficiently replicable and that some difficulties arose in identifying the established criteria. Consequently, it was felt that a careful examination of intra- and inter-observer agreement in assessing these traits would be crucial for a valid estimation of their correct sex allocation accuracies. Although the composite arch has been developed for use in adult remains, its relative success and distinct description justifies testing its accuracy in determining the sex of subadults, and comparing it with the results obtained for the arch criterion.

## 2. Materials and methods

The ilia examined in this study derive from 97 subadult skeletons of known sex and age, housed at the *Bocage* Museum (National Museum of Natural History) in Lisbon Portugal [15]. Several individuals were initially excluded because the ilium was either too fragmented or had already begun to fuse with the ischium and pubis. The age interval of the sample ranges from newborn to 14.99 years of age and a complete age and sex distribution is shown in Table 1. The sex distribution is rather even across the age groups, except between 1.00 and 1.99 years where there is a noticeable excess of males.

Each ilium was laid on a flat surface and an orthogonal photograph was taken. The photographs then became the materials used to assess the composite arch and arch criterion traits. This enabled all observers to examine each trait under exactly the same conditions. The composite arch as defined by Genovés [9], assesses whether the curved line that extends ventrally from the lateral rim of the auricular surface and the curved line that extends dorsally from the anterior rim of the sciatic notch, have the same circumference and overlap each other (single arch = male pattern) or do not overlap (double arch = female pattern) (Fig. 1A). The arch criterion was proposed by Schutkowski [11] and is very similar to the composite arch criterion. It evaluates whether the line that extends cranially (dorsally) from the vertical (anterior) side of the greater sciatic notch leads into the lateral rim of the auricular surface (male) or crosses the auricular surface (female) (Fig. 1B).

The traits were scored separately on each ilium by three different observers, one experienced (observer 1), one inexperienced (observer 2) and one senior osteologist (observer 3). First, unserialied ilia were scored for the arch criterion and, a few weeks later the ilia were serialied for the same trait and scored accordingly. Serialiation is the arrangement of skeletal specimens within a series, from the more male-like to the more female-like individual, forcing the observer to apply relatively consistent scoring standards for the trait being evaluated, thereby reducing error [16]. Although in forensic cases sex determination of specimens is done in isolation without the advantage of arranging skeletal specimens within a series, the purpose here was to assess the benefit of serialiation on accuracy and precision of the traits in sex determination. The composite arch, however, was scored in an unserialied manner, because it is defined in terms of absence/presence and thus it does not allow the observer to sort the ilia in terms of more male-like or more female-like features. Two of the observers also experienced difficulties in allocating some individuals in the more male-like or female-like categories when the arch criterion was scored by serialiation. These individuals were given a score of "undetermined" and were not included in the accuracy tests. Each observer repeated the assessments after an interval of several weeks in order to provide an estimate of intra-observer agreement. The first assessments of each observer were then compared with that of the other observers so as to offer a measure of inter-observer agreement. Percentage of agreement and the  $k$  coefficient were calculated to measure observer concordance.

Table 1  
Age and sex distribution of the study sample ( $N = 97$ )

| Age (years) | Male | Female |
|-------------|------|--------|
| NB–0.99     | 8    | 8      |
| 1.00–1.99   | 15   | 5      |
| 2.00–2.99   | 6    | 2      |
| 3.00–3.99   | 5    | 1      |
| 4.00–4.99   | 3    | 5      |
| 5.00–5.99   | 2    | 2      |
| 6.00–6.99   | 1    | 2      |
| 7.00–7.99   | 2    | 1      |
| 8.00–8.99   | 2    | 2      |
| 9.00–9.99   | 2    | 2      |
| 10.00–10.99 | 2    | 4      |
| 11.00–11.99 | 5    | 3      |
| 12.00–12.99 | 2    | 2      |
| 13.00–13.99 | 1    | 0      |
| 14.00–14.99 | 1    | 1      |
| Total       | 57   | 40     |

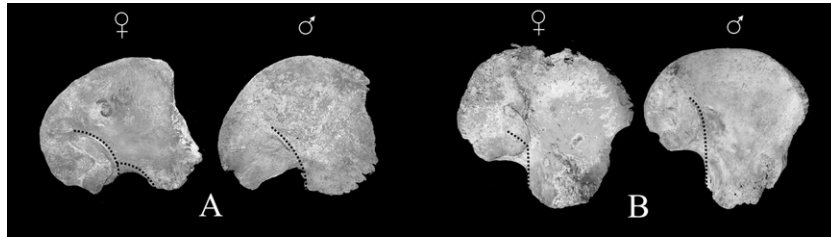


Fig. 1. (A) Illustration of the composite arch as defined by Genovés [9]. The left image depicts the female pattern (double arch) and the right image depicts the male pattern (single arch). (B) Illustration of the arch criterion as defined by Schuttkowski [11]. The left image depicts the female pattern (line crosses auricular surface) and the right image depicts the male pattern (line leads into lateral rim of auricular surface).

### 3. Results

Intra- and inter-observer agreement tests are shown in Table 2. Results show that intra-observer agreement is greater (67.7–88.5%) than inter-observer agreement (50.5–76.3%). However, both traits show an overall poor correspondence between and within observers. In fact, when the  $k$  coefficients are considered only observer 1, when re-scoring the composite arch, obtained a value close to 0.8, which is the cut-off point above which agreement is considered ‘almost perfect’ [17]. Most comparisons only achieved a fair or moderate agreement and three actually only showed slight agreement ( $<0.2$ ) [17]. Poorest agreement, as measured by the  $k$  statistic, was found on inter-observer tests. This is further highlighted by the fact that one-third of inter-observer comparisons did not perform significantly better than chance. Observers 1 and 2 show similar percentages of intra-observer agreement, whereas observer 3 has slightly lower percentages. The highest inter-observer agreement was achieved between observers 1 and 2 and the lowest between observers 2 and 3. Observer error did not show any consistent results between traits, as some tests show higher agreement with the composite arch and others with the arch criterion.

In order to compare the accuracies obtained in this study with those of other studies [11,12], results have been broken down into four age groups. The distribution of correct sex allocation accuracy percentages by age group and observer is presented in Table 3. There is a considerable variation in the association of each trait with sex. Correct sex allocation accuracies range between a low of 0.0% for female expression of the arch criterion in the 11.00–14.99 age group for observers 1 and 2, to a high of 100.0% for the male expression of the composite arch and seriated arch criterion, also in the 11.00–14.99 age group, for observers 2 and 3. When only total correct allocation accuracies are considered, they vary from a low of 23.1% for the seriated arch criterion scored by observer 1 in the

11.00–14.99 age group, to a high of 80.0% for the composite arch scored by observer 2 in the same age group. However, in general, when total allocation accuracies are compared only the composite arch in the 11.00–14.99 age group scored by observers 1 and 2 are close or higher than 75%, the minimum acceptable level of accuracy established by De Vito and Saunders [18]. The traits performed poorly overall and in most cases lead to skewed results, where allocation accuracy is high for one sex but low for the other. The differences in correct allocation accuracy between the sexes were quite large, by as much as 77.8%. Not only did the traits in various age groups only perform slightly better than chance, but occasionally they were also found to show the opposite association with sex. That is, in some cases females scored more consistently as males and vice-versa, as shown by total allocation accuracies under 50%.

The arch criterion and the composite arch do not show significantly different correct sex allocation accuracies, except in the 11.00–14.99 age group, where the composite arch performed better. In addition, seriation of the arch criterion did not improve the overall performance of the trait. There is some variation in allocation accuracies of the three observers, but they performed poorly overall except, again, when the composite arch was scored by observers 1 and 2 in the 11.00–14.99 age group. No trait consistently showed higher accuracies either across observers or age groups. Even the results for the composite arch scored by observer 1 in the 11.00–14.99 age group, which achieved 80% of total correct allocation accuracy, showed considerably skewed results in the two sexes and was not found to perform significantly better than chance, due to the small sample size.

### 4. Discussion

Although morphological traits of the adult pelvis have been described as universally applicable for determining the sex of adults [10,19], some evidence has accumulated for a reduction

Table 2

Intra- and inter-observer error results for both traits, assessed by percentage of agreement and by the  $k$  coefficient (in brackets)

| Trait          | Intra-observer agreement |              |              | Inter-observer agreement |                  |                  |
|----------------|--------------------------|--------------|--------------|--------------------------|------------------|------------------|
|                | Observer 1               | Observer 2   | Observer 3   | Observer 1 vs. 2         | Observer 1 vs. 3 | Observer 2 vs. 3 |
| Arch criterion | 72.9* (0.45)             | 81.3* (0.63) | 67.7* (0.35) | 68.0* (0.36)             | 62.9 (0.26)      | 57.7 (0.15)      |
| Composite arch | 88.5* (0.77)             | 71.8* (0.44) | 70.8* (0.42) | 76.3* (0.53)             | 51.5 (0.03)      | 50.5 (0.01)      |

\*  $p < 0.05$ , being  $p$  the probability that the comparison did not perform significantly better than chance.

Table 3

Percentage of total (%T), female (%F) and male correct (%M) allocation accuracies for each of the traits, assessed by the three observers and broken down by age groups

| Age group (years) | Arch criterion (unseriated) |       |       | Arch criterion (seriated) |       |       | Composite arch |       |       |
|-------------------|-----------------------------|-------|-------|---------------------------|-------|-------|----------------|-------|-------|
|                   | F (%)                       | M (%) | T (%) | F (%)                     | M (%) | T (%) | F (%)          | M (%) | T (%) |
| Observer 1        |                             |       |       |                           |       |       |                |       |       |
| NB–1.99           | 61.5                        | 34.8  | 44.4  | 84.6                      | 26.1  | 47.2  | 84.6           | 39.1  | 55.6  |
| 2.00–5.99         | 40.0                        | 43.8  | 42.3  | 50.0                      | 43.8  | 45.5  | 40.0           | 50.0  | 46.2  |
| 6.00–10.99        | 36.4                        | 77.8  | 55.0  | 20.0                      | 66.7  | 42.1  | 36.4           | 44.4  | 40.0  |
| 11.00–14.99       | 0.0                         | 77.8  | 46.7  | 0.0                       | 42.9  | 23.1  | 66.7           | 77.8  | 73.3  |
| Total             | 40.0                        | 50.9  | 46.4  | 45.7                      | 40.0  | 42.2  | 57.5           | 49.1  | 52.6  |
| Observer 2        |                             |       |       |                           |       |       |                |       |       |
| NB–1.99           | 61.5                        | 21.7  | 36.1  | 62.5                      | 17.6  | 32.0  | 61.5           | 17.4  | 33.3  |
| 2.00–5.99         | 70.0                        | 25.0  | 42.3  | 42.9                      | 41.7  | 42.1  | 30.0           | 56.3  | 46.2  |
| 6.00–10.99        | 45.5                        | 33.3  | 40.0  | 16.7                      | 60.0  | 36.4  | 36.4           | 55.6  | 45.0  |
| 11.00–14.99       | 0.0                         | 66.7  | 40.0  | 0.0                       | 80.0  | 44.4  | 50.0           | 100.0 | 80.0  |
| Total             | 50.0                        | 31.6  | 39.2  | 25.7                      | 27.3  | 26.7  | 45.0           | 47.4  | 46.4  |
| Observer 3        |                             |       |       |                           |       |       |                |       |       |
| NB–1.99           | 61.5                        | 21.7  | 36.1  | 23.1                      | 56.5  | 44.4  | 46.2           | 17.4  | 27.8  |
| 2.00–5.99         | 60.0                        | 37.5  | 46.2  | 50.0                      | 43.8  | 46.2  | 70.0           | 6.3   | 30.8  |
| 6.00–10.99        | 63.6                        | 22.2  | 45.0  | 36.4                      | 77.8  | 55.0  | 54.5           | 33.3  | 45.0  |
| 11.00–14.99       | 50.0                        | 22.2  | 33.3  | 16.7                      | 100.0 | 66.7  | 33.3           | 22.2  | 26.7  |
| Total             | 60.0                        | 26.3  | 40.2  | 32.5                      | 63.5  | 50.5  | 52.5           | 17.5  | 32.0  |

No percentage of total allocation accuracy showed results significantly better than chance, since the *p*-value was always greater than 0.05.

in allocation accuracy when one or a few of these traits are applied to different samples [20–23]. Similarly, this study has shown that the so-called useful traits of the subadult pelvis do not meet expectations of high accuracy when applied to other samples. Compared to Schutkowski [11], results with the arch criterion show a decrease in accuracy of about 15% of more, particularly for males. Reduction in accuracy was even greater when compared to Sutter's [12] test. Moreover, after seriation, accuracy results did not consistently improve, thus further suggesting that the arch criterion is an unreliable trait for determining the sex of subadults in this sample. Although no prior studies have tried to assess the reliability of the composite arch in sex determination of subadult ilia, our results demonstrate that this criterion does not improve the overall accuracy of the greater sciatic notch-auricular surface area for sex determination of the subadult pelvis. Very similar results to this study have also been obtained with other purportedly useful traits for sex determination of subadult skeletons. When the mandibular traits proposed by Loth and Henneberg [24] were independently tested on two other documented collections [25,26] results also showed a very poor overall accuracy, between 41.6% and 63.9%, with significant sex differences in accuracy.

Although Schutkowski [11] obtained a modest correct sex allocation accuracy (72%) with the arch criterion, Sutter's [12] test shows an increased performance (82.3%), suggesting that the trait is useful for sex determination of subadult skeletal remains. However, a closer inspection of Sutter's [12] results reveals a puzzling finding, which may result from the intricacy of sex-related morphologies in subadult ilia. In his study, Sutter [12] considered "bordering" the associated expression of the female trait and "crosses" the expression of the male trait.

However, Schutkowski [11, p. 201] described the *female* pattern as when "the arch (...) crosses the auricular surface" (emphasis added), and the male pattern as when the arch "(...) leads into the lateral rim of the auricular surface". From this, one can only wonder whether there was an error in reporting the results or whether the traits were misapplied. If the second alternative is correct, one is left to conclude that the correct allocation accuracies of the arch criterion are seriously flawed and inflated in Sutter's [12] assessment, thus supporting the results of the current study, which has found this trait as unreliable for sex determination of subadult ilia. More importantly, it may suggest a deficient association between sex and the differential expression of the traits.

Are the results of this study suggesting that the degree of expression of these traits may vary across populations and, therefore, the threshold for establishing criteria for sex determination may also differ from one population to the other? Although specific features can be associated or identified preferably in one of the sexes, is it possible that the threshold for discriminating sex may be towards the "female-like" expression in one population and more towards the "male-like" morphology in another population? At present we cannot tell which factors, besides sex, contribute significantly to the morphology of these features. There could be pathological or morbidity factors as well as individual factors. In fact, it seems there may be considerable variation in greater sciatic notch-auricular surface morphology. This variability is highlighted by the fact that several individuals show an arch criterion pattern that fits with one sex and, at the same time, a composite arch pattern that fits with the opposite sex, particularly the older individuals. Given that both of these traits are defined in relation to the greater sciatic notch-auricular surface area, these

results suggest a poor association of the established criteria with sex in subadults. In addition, specific features are difficult to identify and associate with the sex-specific standards and criteria established by others. The results of the intra- and inter-observer error tests presented here are an expression of such difficulties. Scoring of the traits was difficult to replicate either among or within the three observers, with differing levels of experience. For example, it was frequently difficult to orientate the ilium vertically with the side of the greater sciatic notch to score the arch criterion, or to resolve when the curve passed through the middle of the sciatic notch rather than along the anterior margin of the notch, when scoring the composite arch. In adults, the composite arch has been found to show the poorest agreement, between different observers, of several morphological pelvic traits [27].

Our data suggest that age may be an important factor in the expression of the traits. It is interesting to note that, for individuals over 11 years of age, the composite arch achieved around or close to 80% of correct sex allocation accuracy, when assessed by observers 1 and 2. The fact that this is the age group in which the ilium is closer to adult morphology may explain the higher accuracy obtained with individuals in this age group. Reasons for this may include the attaining of adult shape in the auricular surface and the anterior rim of the sciatic notch becoming more defined. On the other hand, inter-observer agreement results and accuracy tests for the third observer, suggest that the composite arch may be difficult to score, particularly because the iliac spine is absent, making it difficult to draw the circumference. Low reliability of the greater sciatic notch-auricular surface area traits in determining the sex of subadult ilia, may be related to the fact that the development of the sciatic notch shows very high individual variability and that sexual dimorphism in the sciatic notch is only more noticeable from 9 years of age [28].

In humans, the androgen testosterone is first detected in male fetal gonads at about 8 weeks [29]. Testosterone formation increases to a peak plasma concentration at about 16 weeks, with levels comparable to adult males. Subsequently, plasma concentrations decrease to low levels by 24 weeks or the end of the second trimester. In female fetuses, on the other hand, the ovaries only begin synthesizing estrogens and androgens during the late gestational period but the levels are low. It is the increased fetal plasma concentration of testosterone in males between 16 and 24 weeks that accounts for the major sex difference in the gonadal hormones [29]. It has been suggested that these higher fetal male levels of testosterone leads to sufficient dimorphism for sex separation in fetal and infant skeletons [30] but obviously, within such a complex system there may be many other factors that combine to affect the somatic growth of the skeletal system and our ability to detect dimorphism. However, Humphrey's [31] work could provide the answer to why sex-specific traits cannot be seen in the juvenile ilium before puberty. The dimensions of the innominate are among the intermediate-late growing variables of the skeleton, which attain 70% of adult size between ages 6 and 12 years and 90% of adult size only between 12 and 18 years [31]. These results mean that by age 12, the hipbone is

still expected to grow approximately 30%, until adult size is reached. The later expression of sexually dimorphic features in the innominate corresponds with its later development. Early growing parts of the skeleton are generally less sexually dimorphic than later growing elements [31,32] and Humphrey [31] suggests that this results from two factors: (1) sexual dimorphism results of a late divergence of male and female growth, for example, at adolescence, and variables in which growth is complete prior to this divergence would not be expected to show sexual dimorphism; (2) the development of sexual dimorphism may be constrained by the time available for sexual differences to accumulate, i.e. sexual dimorphism is also caused by a difference in male and female growth rates.

Despite the higher fetal male levels of testosterone [29,30] and divergence of male and female post-cranial growth trajectories still prior to adolescence [28,31], because the innominate shows such a late developmental pattern and late attainment of adult size, sexually dimorphic features may not be readily recognizable before puberty. Sexual dimorphism in the greater sciatic notch-auricular surface area, and particular the composite arch, is probably an expression of that developmental trajectory. Another example of the late attainment of sexually dimorphic traits of the innominate is the ventral arc [33]. A precursor ventral arc first appears at age 14 in Sutherland and Suchey's [33] sample, and at age 21 it becomes the most frequent condition. The actual ventral arc is the most frequent condition only by age 23, although there was a case of a young female (age 16), who showed a true ventral arc. In males, there is no ridge in the relevant "arc area", although some males show a distinct line which parallels the symphyseal edge, but only at about age 25. Results obtained by Mittler and Sheridan [34] using the elevation of the auricular surface for sex determination in subadults, also suggest this trait to be age-dependent. In older subadults (>10 years), an elevated articular surface was indicative of the female sex with a virtually 100% probability, although failure to develop a surface elevation was not comparably reliable for males (74.4%) [34].

Although most sexually dimorphic traits may be of limited use in determining the sex of subadults before puberty, additional problems arise from the scoring of morphological methods. These are related to the subjectivity with which each observer identifies traits and discriminates between different expressions of that trait. In comparison, metric methods can provide a more objective and reliable approach [18,35–37], but they may not fulfil the requirements for universal application. Because different situations will require different approaches to sex determination of subadult skeletal remains, morphological criteria should be further explored and evaluated. Special attention, however, should be paid to the description and illustration of the sex-specific traits and to their replicability.

## 5. Conclusion

This study illustrates the difficulty in recognizing and applying morphological traits to subadult ilia and the unreliability of methods when these are applied to different samples, which may vary in the expression of morphological characters. The fact that some adult traits can be identified in

subadult remains by some researchers does not automatically make them useful for sex determination, particularly if established criteria are hard to replicate. Researchers need to be aware of such problems and conduct more blind tests of so-called useful traits by studying samples of known sex, which unfortunately, are hard to come by.

## Acknowledgments

We would like to thank the anonymous reviewers for their helpful comments, Andy Simpson for his participation in the project and Dr. Jaroslav Bruzek for bibliographic information. HC also acknowledges the Bocage Museum administration for institutional support and Fundação para a Ciência e Tecnologia for funding (grant# SFRH/BD/4917/2001 and SFRH/BPD/22142/2005).

## References

- [1] Workshop of European Anthropologists, Recommendations for age and sex diagnoses of skeletons, *J. Hum. Evol.* 9 (1980) 517–549.
- [2] A. Thomson, The sexual differences of the foetal pelvis, *J. Anat. Phys.* 33 (1899) 359.
- [3] D.G. Morton, C.T. Hayden, A comparative study of male and female pelvis in children with a consideration of the etiology of pelvic conformation, *Am. J. Obstet. Gynecol.* 41 (1941) 485–495.
- [4] D.G. Morton, Observations of the development of pelvic conformation, *Am. J. Obstet. Gynecol.* 44 (1942) 789–819.
- [5] E.L. Reynolds, The bony pelvic girdle in early infancy, *Am. J. Phys. Anthropol.* 3 (1945) 321–354.
- [6] E.L. Reynoldes, The bony pelvis in prepuberal childhood, *Am. J. Phys. Anthropol.* 5 (1947) 165–200.
- [7] B.J. Boucher, Sex differences in the fetal sciatic notch, *J. Forensic Med.* 2 (1955) 51–54.
- [8] B.J. Boucher, Sex differences in the fetal pelvis, *Am. J. Phys. Anthropol.* 15 (1957) 581–600.
- [9] S. Genovés, L'estimation des différences sexuelles dans l'os coxal; différences métriques et différences morphologiques, *Bull. Mem. Soc. Anthropol. Paris* 10 (1959) 3–95.
- [10] J. Bruzek, A method for visual determination of sex using the human hip bone, *Am. J. Phys. Anthropol.* 117 (2002) 157–168.
- [11] H. Schutkowski, Sex determination of infant and juvenile skeletons, I—morphognostic features, *Am. J. Phys. Anthropol.* 90 (1993) 199–205.
- [12] R.C. Sutter, Nonmetric subadult skeletal sexing traits: I—A blind test of the accuracy of eight previously proposed methods using prehistoric known-sex mummies from northern Chile, *J. Forensic Sci.* 48 (2003) 927–935.
- [13] V. Novotný, Pohlavní rozdíly a identifikace pohlaví pánevní kosti [Sex differences and identification of sex in pelvic bone], Ph.D. Thesis, Purkyně University, Brno, 1981.
- [14] L. Klepinger, *Fundamentals of Forensic Anthropology*, John Wiley and Sons Inc., New York, 2006.
- [15] H.F.V. Cardoso, The collection of identified human skeletons housed at the Bocage Museum (National Museum of Natural History) in Lisbon, Portugal, *Am. J. Phys. Anthropol.* 129 (2006) 173–176.
- [16] C.O. Lovejoy, Dental wear in the Libben population: its functional pattern and role in the determination of adult skeletal age at death, *Am. J. Phys. Anthropol.* 68 (1985) 46–56.
- [17] H.C. Kraemer, V.S. Periyakoi, A. Noda, Kappa coefficients in medical research, in: R.B. D'Agostino (Ed.), *Tutorials in Biostatistics, Statistical Methods in Clinical Studies*, vol. 1, John Wiley & Sons, New York, 2004 pp. 85–105.
- [18] C. De Vito, S.R. Saunders, A discriminant function analysis of deciduous teeth to determine sex, *J. Forensic Sci.* 35 (1990) 845–858.
- [19] J. Dittrick, J.M. Suchey, Sex determination of prehistoric central California skeletal remains using discriminant analysis, *Am. J. Phys. Anthropol.* 70 (1986) 8–9.
- [20] N.C. Lovell, Test of Phenice's technique for determining sex from the os pubis, *Am. J. Phys. Anthropol.* 79 (1989) 117–120.
- [21] S.M. MacLaughlin, M.F. Bruce, The accuracy of sex identification in European skeletal remains using the Phenice characters, *J. Forensic Sci.* 35 (1990) 1384–1392.
- [22] T. Rogers, S.R. Saunders, Accuracy of sex determination using morphological traits of the human pelvis, *J. Forensic Sci.* 39 (1994) 1047–1056.
- [23] J. Albanese, A metric method for sex determination using the hipbone and femur, *J. Forensic Sci.* 48 (2003) 263–273.
- [24] S.R. Loth, M.B. Henneberg, Sexually dimorphic mandibular morphology in the first few years of life, *Am. J. Phys. Anthropol.* 115 (2001) 179–186.
- [25] L.B. Scheuer, Brief communication: a blind test of mandibular morphology for sexing mandibles in the first few years of life, *Am. J. Phys. Anthropol.* 119 (2002) 189–191.
- [26] H. Coqueugnot, G. Giacobini, G. Malerba, L'utilisation de caractères morphologiques dans la diagnose sexuelle des mandibules d'enfants: application à la collection ostéologique de Turin (Italie), *Bull. Mem. Soc. Anthropol. Paris* 14 (2002) 131–139.
- [27] M. Đurić, Z. Rakočević, D. Đonic, The reliability of sex determination of skeletons from forensic context in the Balkans, *Forensic Sci. Int.* 147 (2005) 159–164.
- [28] W.H. Coleman, Sex differences in the growth of the human bony pelvis, *Am. J. Phys. Anthropol.* 31 (1969) 125–152.
- [29] M.M. Grumbach, F.A. Conte, Disorders of sex differentiation, in: J.D. Wilson, D.W. Foster (Eds.), *Williams Textbook of Endocrinology*, WB Saunders Co., Philadelphia, 1992, pp. 1303–1425.
- [30] S.R. Saunders, Subadult skeletons and growth related studies, in: S.R. Saunders, M.A. Katzenberg (Eds.), *Skeletal Biology of Past Peoples: Research Methods*, Wiley-Liss, New York, 1992, pp. 1–20.
- [31] L.T. Humphrey, Growth patterns in the modern human skeleton, *Am. J. Phys. Anthropol.* 105 (1998) 57–72.
- [32] A.H. Schultz, Metric age changes and sex differences in primate skulls, *Z. Morphol. Anthropol.* 52 (1962) 239–255.
- [33] L.D. Sutherland, J.M. Suchey, Use of the ventral arc in pubic sex determination, *J. Forensic Sci.* 36 (1991) 501–511.
- [34] D.M. Mittler, S.G. Sheridan, Sex determination in subadults using auricular surface morphology: a forensic science perspective, *J. Forensic Sci.* 37 (1992) 1068–1075.
- [35] F.W. Rösing, Sexing immature human skeletons, *J. Hum. Evol.* 12 (1983) 149–155.
- [36] M. Teschler-Nicola, H. Prossinger, Sex determination using tooth dimensions, in: K. Alt, F.W. Rösing, M. Teschler-Nicola (Eds.), *Dental Anthropology—Fundamentals, Limits, and Prospects*, Springer Verlag, Wien, 1998, pp. 479–500.
- [37] H.F.V. Cardoso, Sample-specific (universal) metric approaches for determining the sex of immature human skeletal remains using permanent tooth dimensions, *J. Archaeol. Sci.* 35 (2008) 158–168.