An assessment of the repeatability of pubic and ischial measurements

M Steyn[1], PJ Becker[2], EN L'Abbé[1], Y Scholtz[1] and J Myburgh[1]

[1]Forensic Anthropology Research Centre, Department of Anatomy, University of Pretoria, South Africa

[2]Medical Research Council, Pretoria, South Africa & Department of Anatomy, University of Pretoria, South Africa

Number of pages: 13

Number of figures: 3

Number of tables: 2

Corresponding author: Prof M Steyn, Department of Anatomy, PO Box 2034, Pretoria 0001, South Africa

e-mail: maryna.steyn@up.ac.za

Tel: 27-12-4203256

ABSTRACT

Forensic anthropologists frequently use measurements of the human skeleton to determine sex and ancestry. Since the establishment of the Daubert criteria of admissibility of scientific evidence to court, methodologies used by anthropologists came under severe scrutiny. It is therefore important to ensure that the osteometric standards that are used in skeletal analyses are clearly explained, repeatable and reliable. Adams and Byrd (2002) found that measurements of the pelvis that originated from a point inside the acetabulum could not be repeated accurately. The purpose of this paper was to use three different sets of pubic and ischial measurements to establish whether they can be repeated with high precision between four different observers, and also by the observers themselves. Generally, high levels of repeatability were obtained, with intra-class correlations (ICC) above 0.8. Pubic and ischial measurements using a point in the acetabulum as origin performed the worst (ICC values of 0.82 and 0.79 respectively for inter-observer repeatability), whereas other methods performed better with ICC values above 0.9. It is advised that pubic and ischial measurements should be taken using the origin of the iliac blade as landmark.

KEYWORDS: forensic anthropology, os coxa, pubis length, ischium length, sex determination, osteometry

Forensic anthropologists are continually attempting to find new methods to use in skeletal identification, or are re-evaluating and refining existing methods. Methods to determine the sex and ancestry of unknown skeletons are either based on morphological (non-metric) or metric characteristics of the various bones of the human body. It is of the utmost importance that these methods (both metric and non-metric) are accurate, precise and repeatable.

Due to the fact that forensic anthropologists can be requested to defend their results in a court of law, the anthropological methodologies which they apply to any case of forensic origin should be continually scrutinized for scientific validity and applicability in the population. A solid scientific approach to skeletal analysis is paramount not only to their professional career, but also to the accurate identification of the victim and, if needs be, conviction of the criminal involved.

In the United States, 1975 Federal Rules of Evidence (FRE) Rule 702 states that the trial judge is in charge of deciding the relevance and reliability of expert testimony [1]. In order to fulfill these duties, they must evaluate the methodology of the research according to the Daubert standards, which were named after the well-known court case of Daubert vs. Merrel Dow Pharmaceuticals [1, 2, 3]. The Daubert guidelines place emphasis on the methods used for analysis (as well as the testability, repeatability and validity), rather than the experience of the person providing the testimony. This has led to a shift in defining the goals of forensic anthropological research, the collection of evidence and the overall analysis of skeletal remains [2, 3]. In essence, the connection between the data and the methods used is extremely important, and the expertise, alone, of any specialist is no longer adequate to ensure acceptance in a court of law.

While the American judicial system does not influence South African law, the outlined regulations for the acceptance of expert testimony provides all forensic anthropologists with much introspection into the various anthropological techniques and methods which have been used and presented as expert testimony in the country, as well as the criteria necessary for one to engage in the practice of being a forensic anthropologist.

The testing and replication of methods have become more important. This also pertains to our ability to precisely and repeatedly measure elements of the human skeleton, as many of our methods to determine sex, for example, rely on the ability to accurately measure skeletal elements. This was the focus of a study by Adams & Byrd [4], who tested the interobserver repeatability of selected post-cranial measurements. These authors concluded that pubic length (amongst others) could not be reliably measured. They measured it from the point in the acetabulum where the three elements in the innominate meet, to the upper end of the pubic symphysis. These authors advised that this measurement should not be used at all, and by implication also not the ischial length as it uses the same point of origin which is difficult to locate.

A large volume of studies have been published where various measurements of the os coxa have been used to assess sexual dimorphism [5-13]. Although pubic length reflects the differences between the sexes very well, this begs the question as to which, if any, of the measurements reflecting pubic and ischial length should be used in metric determination of sex from the pelvis. Metric estimation of sex from the pelvis (especially discriminant function analysis), becomes particularly important when dealing with incomplete or fragmented remains, or cases where the morphology is ambiguous [8, 9,

11, 12, 13]. This is of special importance in South African individuals as methods like the Phenice method have not yet been tested on South African populations.

The aim of this study was to measure pubic and ischial length, using three different methods described in the literature, in order to assess which of them (if any) can be measured reliably and with a high degree of repeatability. For this purpose four observers measured and remeasured the same sample of os coxae, and the repeatability was assessed.

Materials and Methods

The left os coxae of 100 individuals from the Pretoria Bone Collection, housed in the Department of Anatomy, University of Pretoria, were used [14]. This collection is based on cadavers of donors and unclaimed persons from nearby hospitals that were used in the dissection halls of the medical school prior to entering the skeletal collection. The sample includes people from various socioeconomic groups [14].

The sample was selected to include 25 undamaged os coxae of each sex-ancestral group represented in the collection (black males, white males, black females and white females). This was simply done to ensure that all possible variation is included in the sample to be measured, because as Adams and Byrd [4] pointed out, the error rate of a method using measurements not only depends on error rates due to inconsistent measurement, but also the natural variability in the trait being measured.

Four observers took a set of six measurements from each left os coxa, which they then repeated a second time (at a later stage). The four observers had varying degrees of experience. Two anthropologists have PhD's (observers A and B) and two were post-

graduates with some experience in measuring skeletal elements (observers C and D). Observers A and B have both authored and co-authored several papers on physical and forensic anthropology and have examined more than 400 cases of a forensic nature. Observers C and D both have masters degrees in Anatomy with specialization in physical anthropology and both have assisted observers A and B in the analysis of various of the above-mentioned cases. The same electronic caliper was used by all observers throughout the study. In order to ensure good comprehension of the landmarks that were to be used, the four observers met before the project commenced and discussed the details of the measurements. No one was allowed to make any markings on any of the bones, and all measurements were taken to one tenth of a millimeter.

The following measurements were recorded for pubis length:

Pubis length a

This measurement was taken from the point in the acetabulum where the ilium, ischium and pubis meet, to the medial border of the pubic symphysis [15]. According to Adams and Byrd [4], the measuring point in the acetabulum may be identified because 1) there may be an irregularity there, both in the acetabulum and inside the pelvis; 2) a change in thickness may be observed when holding the bone up to a light; 3) frequently a notch can be observed in the border of the articular surface of the acetabulum. In measuring the pubis, care should be taken to hold the caliper parallel to the long axis of the bone. This is the same measurement as was used in the study by Adams & Byrd [4]. Consensus was reached by the four observers to record this measurement from a point close to or on the edge of the smooth articular (lunate) surface inside the acetabulum (Fig. 1), meeting one or all of the criteria as outlined above.

Pubis length b

This measurement was recorded from the medial border of the pubic symphysis to the point where the iliac blade meets the acetabulum. This landmark is defined as the point on the superior border of the acetabulum at the center of the origin of the iliac blade, and was used by Patriquin *et al.* [8] to circumvent the problems with using the point inside the acetabulum (Fig. 2).

Pubis length c

This dimension is measured from the medial border of the pubic symphysis to the closest point on the rim of the acetabulum [11,12] (Fig. 3).

The following measurements were recorded for ischium length, and the landmarks correspond to those used for pubis length:

Ischial length a

This measurement is taken from the point in the acetabulum where the ilium, ischium and pubis meet as described above, to the distal end of the ischium [4] (Fig. 1).

Ischial length b

This is measured from the point on the superior ridge of the acetabulum at the center of the origin of the iliac blade as described above, to the deepest point on the ischial tuberosity [8] (Fig. 2).

Ischial length c

This dimension is measured from the most distal point of the ischium to the closest point on the rim of the acetabulum [11, 12] (Fig. 3).

The repeatability/reproducibility of the measurements was assessed using the intra-class correlation coefficient (ICC) [16], which is calculated from the between and within subject sums of squares in a one-way analysis of variance, using ordinary least squares estimation (LSE). An ICC value of 1 indicates 100% repeatability. Thus the closer the values are to one, the higher the repeatability. The ICC measures the repeatability between the four observers (inter-observer repeatability) when their first readings only are analyzed together, while when the two measurements by the same observer is analyzed within the observer the ICC measures the repeatability within the observer (intra-observer repeatability).

The means, median and 95% confidence interval for each of the measurements, as scored by each of the observers, were also calculated. This shows the ability of each of the observers to record the measurements accurately, in order to assess if the variability observed is particularly due to one individual or not. In addition it also gives an estimate, through the 95% confidence interval (CI), of how big the differences in actual dimensions are. Stata Release 10 software was used.

Results

The results for inter-observer repeatability are shown in Table 1. Both measurements using the point inside the acetabulum as a landmark (a) show moderate to good repeatability. For the pubis measurement, a 0.826 agreement rate was obtained between the four observers, while for the ischium it was less at 0.795.

Measurement (b), where the origin of the iliac blade was used as landmark fared the best, with ICC values of 0.912 for the pubis and 0.952 for the ischium length,

respectively. The measurements to the closest part of the acetabular rim (c) showed

repeatabilities of 0.915 and 0.912 respectively for the pubis and ischium.

The closeness of the mean and median values for each measurement shows that

the data have a normal distribution (Table 2). From these values it can be seen that, in

general, all four observers obtained mean values which were very close to each other and

differed by one or two millimeters only. The only possible exception here is ischial

length a, where Observer B obtained values somewhat higher than those of the other

three observers.

Intra-observer repeatabilities for the four observers (A-D) are shown in Table 3.

For pubis length (a) the ICC ranged from 0.872 to 0.987, which, as can be expected, is

higher than what was observed between observers. Ischial length (a) had intra-observer

repeatabilities ranging from 0.929 - 0.971. Pubis length (b) showed ICC's ranging

between 0.935 and 0.979, generally indicating higher consistency than pubis length (a).

The corresponding ischial measurements (ischium length b) had ICC values of 0.939 –

0.988. Intra-class correlation values for the third set of measurements ranged from 0.904

to 0.987 for pubis length (c), and 0.953 - 0.981 for ischium length (c).

Discussion

Although not directly comparable due to the use of different statistical methods, the

results from this study are generally in agreement with that of Adams & Byrd [4] who

demonstrated that measurements which used the point inside the acetabulum could not be

measured consistently between observers. The measurements using this landmark only

had ICC values ranging between 0.79 and 0.82, which may not be good enough [17].

Generally speaking an ICC value of 0.9 indicates excellent repeatability, although values above 0.8 are also good. These landmarks should therefore preferably not be used in any metric analysis.

On the positive side, both other methods using different landmarks performed better, with ICC values above 0.9. The measurements using the origin of the iliac blade as landmark, although not perfectly repeatable, show good consistency. In addition, especially the ischium length dimension (ischium length b) is good to elucidate sexual dimorphism, as it also includes the full width of the acetabulum and can thus be expected to be much larger in males. On the downside, the inclusion of part of the (larger) acetabulum in males in the pubis length measurement somewhat obscures the dimorphism, as it masks the longer pubis length expected in females [7, 18].

As expected, the intra-observer repeatability is higher than the inter-observer repeatability, and it is possible that each observer had a preconceived notion of where each landmark should be and thus managed to find high repeatability. The fact that some of the values obtained by observer B differed from the rest of the observers (especially ischial length b) illustrate the importance of calculating inter-observer repeatability.

It seems that in this case experience did not make much of a difference, as the less experienced and more experienced observers performed on an equal level. This may have been due to the fact that all observers met and discussed the landmarks before the project commenced, showing that clear definition of landmarks and measurements and good preparation is essential to produce acceptable results. In this regard this study differs from that of Adams & Byrd [4] who asked many observers to measure bones, based on descriptions in the literature only. It shows that a certain degree of training and

preparation may be necessary for more complex measurements to ensure that results are repeatable.

In conclusion, this study confirms the observation by Adams and Byrd [4] that pubic and ischial length measurements using the landmark in the acetabulum where the three elements fuse, cannot be measured very reliably although some of the problems with this landmark can be avoided with careful planning and training of observers. However, there are two other methods of measuring the same pelvic elements, and these can be used with good results. Particularly the measurement using the origin of the iliac blade [8] showed good repeatability and it is advised that this measurement should be used in any study where pelvic measurements are required.

Acknowledgements

References

[1] A.M. Christensen, C.M. Crowder, Evidentiary standards for forensic anthropology, J. Forensic Sci. 54 (2009) 1211 – 1216.

[2] D.C. Dirkmaat, L.L. Cabo, S.D. Ousley, S.A. Symes, New perspectives in forensic anthropology, Am. J. Phys. Anthropol. Suppl 47 (2008) 33 – 52.

[3] C.R. Grivas, D.A. Komar, *Kumho, Daubert,* and the nature of scientific inquiry: implications for forensic anthropology, J. Forensic Sci. 53 (2008) 771 – 776.

[4] A.J. Adams, J.E. Byrd, Interobserver variation of selected postcranial measurements, J. Forensic Sci. 47 (2002) 1193 – 1202.

[5] R.G. Tague, Variation in pelvic size between males and females, Am. J. Phys. Anthropol. 80 (1989) 59 – 71.

[6] V. Davivongs, The pelvic girdle of the Australian Aborigine; sex differences and sex determination, Am. J. Phys. Anthropol. 21 (1963) 443 – 445.

[7] S.M. MacLaughlin, M.F. Bruce, Sex determination from the pelvis in a Dutch skeletal series, J. Anat. 140 (1985) 532.

[8] M.L. Patriquin, M. Steyn, S.R. Loth, Metric analysis of sex differences in South African black and white pelves, Forensic Sci. Int. 147 (2005) 119 – 127.

[9] M.H. Day, R.W. Pitcher-Wilmott, Sexual differentiation in the innominate bone studied by multivariate analysis, Ann. Hum. Biol. 2 (1975) 143 – 151.

[10] C. Rissech, M. Garcia, A. Malgosa, Sex and age diagnosis by ischium morphometric analysis, Forensic Sci. Int. 135 (2003) 188 – 196.

[11] F.P. Schulter-Ellis, L.C. Hayek, D.J. Schmidt, Determination of sex with a discriminant analysis of new pelvic bone measurements: Part II, Forensic Sci. Int. 30 (1985) 178 – 185.

[12] M. Steyn, M.Y. İşcan, Metric sex determination from the pelvis in modern Greeks, Forensic Sci. Int. 179 (2008) 86.e1 – 6.

[13] M. Steyn, M.L. Patriquin, Osteometric sex determination – Does population specificity matter? Forensic Sci. Int. 191 (2009) 113.e1 – 5.

[14] E.N. L'Abbé, M. Loots, J.H. Meiring, The Pretoria Bone Collection: a modern South African sample, Homo 56 (2005) 197 – 205.

[15] J.E. Buikstra, D.H. Ubelaker, Standards for data collection from human skeletal remains, Arkansas Archeological Survey, Fayetteville, 1994.

[16] J.M. Lachin, The role of measurement reliability in clinical trials, Clin. Trials 1 (2004) 553 – 566.

[17] L. Ferrante, L. Cameriere, Statistical methods to assess the reliability of measurements in the procedures for forensic age estimation, Int. J. Legal Med. 123 (2009) 277 – 283.

[18] K.R. Nagesh, T. Kanchan, B.K. Bastia, Sexual dimorphism of acetabulum-pubis index in South-Indian population, Leg. Med. 9 (2007) 305 – 308.

Table 1: ICC between observers A-D

	ICC – least squares	95% confidence interval
Pubis length a	0.826	(0.777;0.874)
Pubis length b	0.912	(0.886;0.939)
Pubis length c	0.915	(0.890;0.941)
Ischium length a	0.795	(0.739;0.850)
Ischium length b	0.952	(0.938;0.967)
Ischium length c	0.912	(0.885;0.938)

Table 2: Measurement values with means and medians for observers A – D .

Observer		Pub a	Pub b	Pub c	Isch a	Isch b	Isch c
A	Mean	77.8	94.1	70.3	86.6	101.3	53.2
	Median	77.3	94.0	69.9	86.1	100.5	52.6
B	Mean	78.4	95.8	68.9	90.9	104.1	53.9
	Median	77.9	95.3	68.1	90.4	102.9	53.0
C	Mean	79.3	95.6	69.7	85.8	101.9	52.1
	Median	78.1	94.7	69.9	85.4	100.5	51.7
D	Mean	76.8	93.7	69.5	85.1	101.4	52.3
	Median	76.5	93.1	69.2	84.9	100.7	51.5

Pub a = pubis length a, Pub b = Pubis length b; Pub c = pubis length c; Isch a = ischial length a; Isch b = ischial length b; Isch c = ischial length c

Table 3: ICC within observers A – D (intra-observer repeatability).

Obs		Pub a	Pub b	Pub c	Isch a	Isch b	Isch c
A	ICC	0.9084	0.955	0.968	0.971	0.939	0.981
	95% CI	(0.870;0.940)	(0.938;0.973)	(0.956;0.981)	(0.960;0.982)	(0.916;0.962)	(0.974;0.988)
B	ICC	0.872	0.935	0.987	0.929	0.9750	0.971
	95% CI	(0.825;0.919)	(0.911;0.960)	(0.982;0.992)	(0.902;0.956)	(0.965;0.985)	(0.960;0.982)
C	ICC	0.896	0.952	0.983	0.936	0.977	0.953
	95% CI	(0.858;0.935)	(0.934;0.970)	(0.976;0.990)	(0.912;0.960)	(0.968;0.986)	(0.935;0.971)
D	ICC	0.987	0.979	0.904	0.952	0.988	0.974
	95% CI	(0.982;0.992)	(0.970;0.987)	(0.868;0.940)	(0.934;0.971)	(0.983;0.993)	(0.964;0.984)

Pub a = pubis length a, Pub b = Pubis length b; Pub c = pubis length c; Isch a = ischial length a; Isch b = ischial length b; Isch c = ischial length c; 95% CI = 95% confidence interval; ICC = Intra-class correlation coefficient
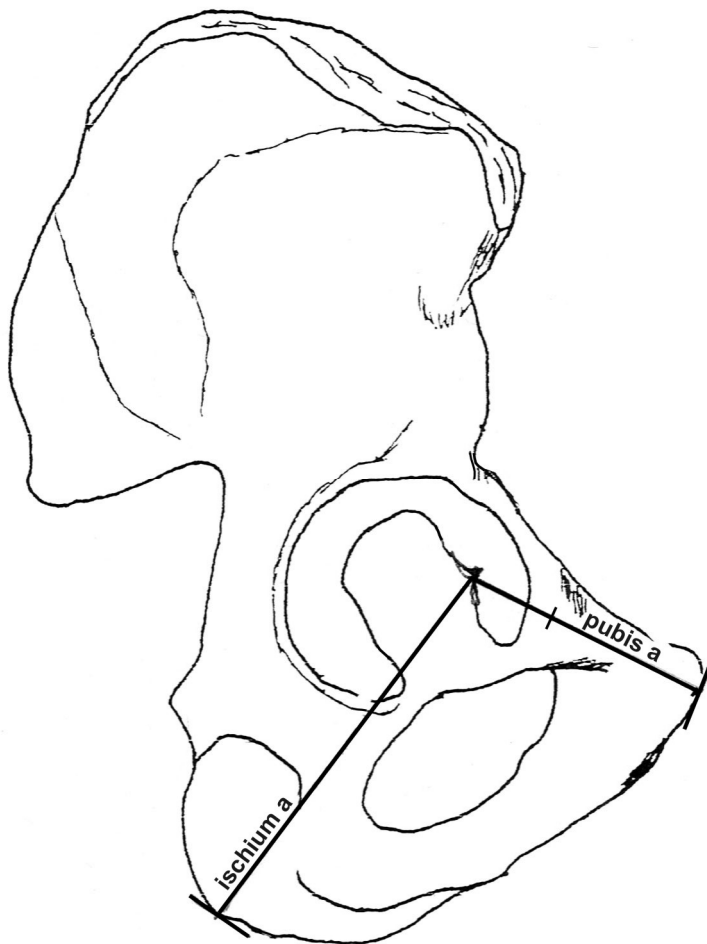
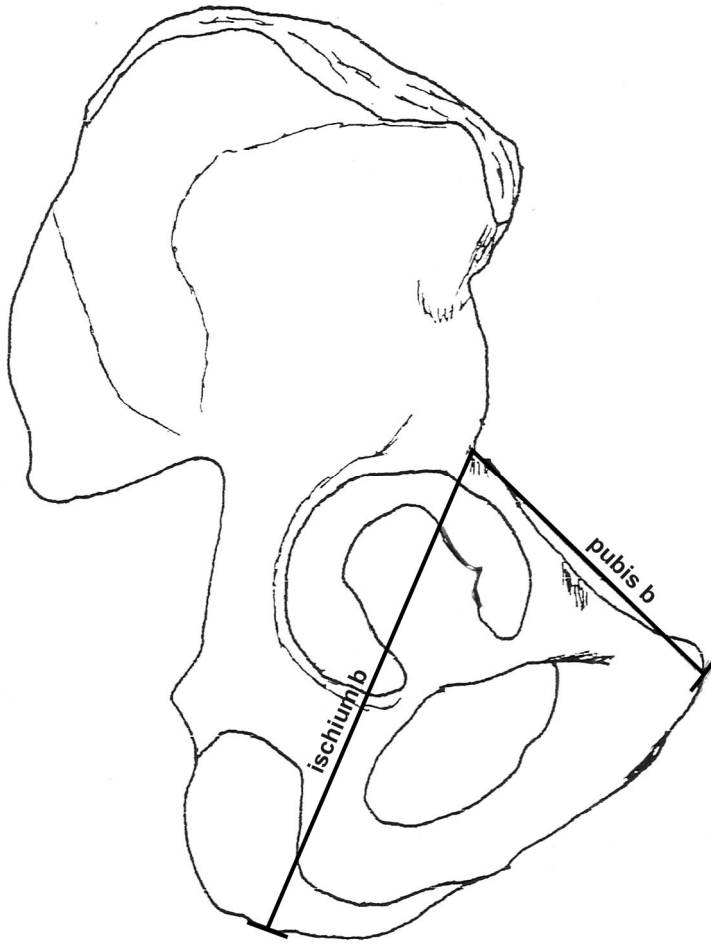**Fig. 1** Measurement of pubis length a and ischium length a
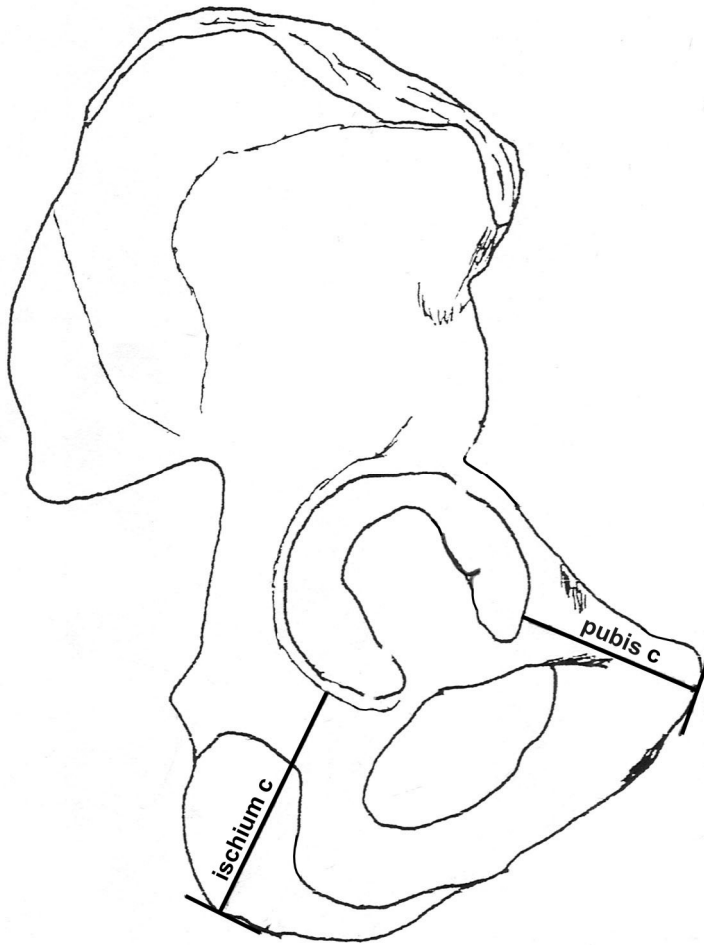
Fig. 2 Measurement of pubis length b and ischium length b

Fig. 3 Measurement of pubis length c and ischium length c