

Congruence of Methods for Determination of Sex using Real, Virtual and 3-D Printed Specimens

Julia Gamble

University of Manitoba, Canada. julia.gamble@gmail.com

Amanda Blackburn

University of Manitoba, Canada. umblack4@cc.umanitoba.ca

Robert D. Hoppa

University of Manitoba, Canada. hoppard@cc.umanitoba.ca

Abstract:

Increasingly, physical anthropology is seeing the use of a variety of digital technologies to capture, describe and analyse skeletal elements. The last ten years have seen a dramatic increase in the number of publications undertaking validations of osteological techniques using CT (computed tomography) data and/or virtual models. In the last few years, with the increasing availability of relatively low cost technological approaches to rapid prototyping, especially 3D printing (3DP), production of bone replicas has gained popularity. However, no studies have explored the relative congruence of methods on real, virtual and 3DP models. This paper presents the results of a study to explore congruency in metric and morphological methods for determination of sex from the human hipbone. Intra- and inter-observer agreement between real, 3D virtual models and 3DP models is assessed using the Kappa statistic. Implications for future osteological studies are discussed.

Key Words: *Physical Anthropology, Sex Determination, 3D Models, Rapid Prototyping*

Introduction

Determination of sex is a central component to any osteological analysis, as it is essential for everything from age estimation to the reconstruction of palaeodemographic profiles and the study of patterns in past population health. Various techniques, including both metric and nonmetric, have been developed for determination of sex from the skeleton. More traditional methods utilize characteristics of the skull and pelvis and studies continue to develop new methods for these as well as other postcranial elements.

Visual analysis of gross morphology has played a central role in the development of techniques

in sex estimation and is still the predominant form of analysis for a number of reasons (Walker 2008). One of the primary benefits of visual methods is the speed with which they can be applied to attain results (Walker 2008). Furthermore, visual methods generally require little to no equipment and can therefore be conducted in almost any situation. They are non-destructive and a number of techniques (particularly those associated with the pelvis) have a demonstrated high level of accuracy.

Bruzek and Murail (2006) identify three morphofunctional segments of the pelvis relevant to understanding sexual dimorphism – the sacroiliac segment, the ischiopubic segment, and the acetabular segment. The

siatic notch is the most important element for sex determination in the sacroiliac segment, while the subpubic angle and the relative lengths of the pubis and ischium are connected with the ischiopubic segment. The form of the third segment is primarily linked with the “spatial organization of the three bones that form the pelvis and contributes to its general architecture” (Bruzek and Murail 2006, 228). Methods developed by Phenice (1969) for sex estimation from the pelvis have been hugely influential in subsequent studies. Phenice defined three pelvic morphological features which were sexually dimorphic: the ventral arc, the subpubic concavity, and the medial aspect of the ischio-pubic ramus. The accuracy of techniques developed from the pelvis has been well established, resulting in their widespread use in sex estimation of skeletal material (see for example Rogers and Saunders 1994; Steyn and Patriquin 2009; Ubelaker and Volk 2002). The development of well-defined techniques can significantly reduce inter-observer error and increase reliability (Pietrusewsky 2008; Walrath et al. 2004). Those techniques that rely on shape rather than size information tend to be more reliable in comparisons across populations (Bruzek and Murail 2006; Buikstra and Ubelaker 1994; Ferembach et al. 1980; Walrath et al. 2004).

The possibility of providing a continuous set of values makes metric analyses particularly useful for applying multivariate statistical techniques (Pietrusewsky 2008). However, difficulties with metric analyses include how to quantify visually identifiable sexually dimorphic features. Often morphological aspects which have demonstrated high levels of accuracy in the past are difficult to capture metrically (Walker 2008). Furthermore, objectivity is not necessarily any easier to achieve metrically than it is visually, as metric techniques require clearly defined and easy to identify points of measurement (Pietrusewsky 2008).

More and more, physical anthropological

research is using various digital technologies to acquire, analyse and interpret osteological data. The past decade has seen a dramatic increase in the number of studies using CT or laser scan data and/or virtual models. More recently, with the increasing availability of relatively low cost technological approaches to rapid prototyping (3DP), physical reconstructions of 3D models have also gained popularity. Studies in the 1990s recognized the potential of stereolithography for anthropological applications (e.g. Hjalgrim et al. 1995; Weber 2001), but the major focus remained on medical applications and the accuracy of models produced using this technology (e.g. Barker et al. 1994; Bouyssie et al. 1997; Choi et al. 2002). Ultimately, the time and expense of producing scans and stereolithographic models remained a barrier to wide-scale use within anthropology. Over the past decade, however, improvements in technology (both in scanning and rapid-prototyping) have resulted in the increased use of CT, laser scanning and 3D printing (3DP) technologies for anthropological research. Most studies assessing these technologies have presented case studies (e.g. Allard et al. 2005; Fantini et al. 2008) or focused on the replicability of 3D landmarking using different data acquisition techniques (e.g. Park et al. 2006; Sholts et al. 2010, 2011), though, Decker et al. (2011) have assessed sex determination from 3D virtual models produced from CT data. However, no papers have explored the relative congruence of methods on real, virtual and 3DP models. This paper presents the results of a study to explore congruence in metric and morphological methods for determination of sex from the pelvis across these modalities.

Materials and Methods

The validity of any osteological technique can be measured in its level of accuracy, precision, and replicability. It is essential for a method to estimate sex with high levels of accuracy and precision through the use of clearly defined features which can be recognized and

consistently replicated by other observers. Most validation studies of methods for sex determination therefore focus on issues of accuracy and precision including a rigorous analysis of intra- and inter-observer error in assessment. A variety of factors can lead to reduced success or high levels of error using a specific method. With the increasing use of multiple modalities (e.g. dry bone specimens, CT data, laser scan data, 3DP specimens) on which to undertake analysis, it becomes critical to establish the impact of modality on the precision of methods. This has most frequently been explored with respect to craniometrics and 3D landmarking of the skull (e.g. Park et al. 2006; Sholts et al. 2010, 2011; Williams

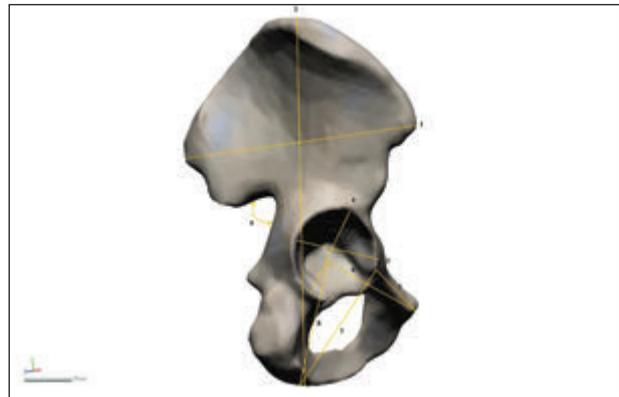


Figure 1. Schematic representation of the measurements taken in the current study (see Table 1 for key).

Feature	Reference
Greater Sciatic Notch	Buikstra and Ubelaker (1994)
Ventral Arc	Phenice (1969); Sutherland and Suchey (1991)
Subpubic Concavity	Phenice (1969)
Medial Aspect of the Ischiopubic Ramus	Phenice (1969)
Preauricular Sulcus	France (1998); Schwartz (1995)
Acetabulum	France (1998); Schwartz (1995)
Obturator Foramen	France (1998); Schwartz (1995)
1. Max. Iliac Breadth	Arsuaga and Carretero (1994)
2. Max. Coxal Bone Length	Arsuaga and Carretero (1994)
3. Transverse Acetabular Diameter	Arsuaga and Carretero (1994)
4. Vertical Acetabular Diameter	Arsuaga and Carretero (1994)
5. Superior Pubic Ramus Length	Albanese (2003)
6. Pubic Length	Schwartz (1995)
7. Acetabulum-Ischium Length	Albanese (2003)
8. Ischial Length	Schwartz (1995)
9. Sciatic Notch Angle	France (1998)
10. Sciatic Notch Width	Schwartz (1995)

Table 1. List of morphological traits and measurements.

and Richtsmeier 2003) but has not yet been substantially investigated systematically for osteological methods of personal reconstruction such as sex determination.

Sex was assessed on a sample of 29 human hipbones using standard morphological and metric techniques. The purpose was not to test the accuracy of the methods being used (since true sex is unknown) but rather the replicability of each trait or measure between observers and across modalities. 3D scan data were created using a NextEngine 3D tabletop colour scanner (accuracy of 0.38mm in wide mode) and further processed using INUS Rapidform reverse engineering software. A subsample of 7 randomly selected models were converted to STL files and printed using a Stratus Dimension 1200es 3D printer (layer thickness of 0.254mm). All analyses were undertaken in the Bioanthropology Digital Image Analysis Laboratory in the Department of Anthropology at the University of Manitoba.

A variety of traditional visual methods and measurements (Fig. 1 and Table 1) were collected on each of the real bone specimens and the 3DP models. Measurements were collected using a Mitutoya digital sliding calliper. These were also collected on the virtual models by manipulating the virtual model in 3D space and by using the measurement tools available

in Rapidform. Assessments of inter- and intra- observer error in congruence of determination of sex from multiple traits was assessed across each modality.

Results

For each specimen, all traits and measures were collected and an assessment of sex based on each individual trait or measure was made. Congruence between multiple traits or measurements within a single specimen was assessed by calculating the percentage agreement across the range of morphological traits or measurements with respect to final determination of sex. A value of 50% reflects half of the traits being in agreement while a value of 100% reflects all of the traits being in agreement. The mean level of congruence for morphological traits was 92% and 78-79% for the metrics in real and virtual specimens respectively.

A final morphological and metric assessment of sex was made for each specimen. Comparison of the final assessment of sex for each of the modalities was then examined. Overall, results from real, virtual, and 3DP pelvises demonstrated high levels of congruence or the same final determination of sex between modalities (Table 2).

		Virtual							
		Morphology				Metrics*			
		M	F	M	F	M	F	M	F
Real	M	22	0	19	1				
	F	0	7	0	7				
		3DP							
		Morphology				Metrics			
		M	F	M	F	M	F	M	F
Real	M	5	0	4	1				
	F	0	2	0	2				

Table 2. Overall congruency of final sex determination for each modality (Observer 1).

Trait	Real Specimens		Virtual Specimens	
	Kappa	p	Kappa	p
Greater Sciatic Notch	1.000	0.001	0.847	0.002
Ventral Arc	0.751	<0.001	1.000	<0.001
Subpubic Concavity	1.000	<0.001	0.914	<0.001
Medial Aspect of Ischiopubic Ramus	0.519	0.005	0.796	<0.001
Preauricular Sulcus	0.650	0.001	0.867	<0.001
Acetabulum	1.000	<0.001	0.576	0.001
Obturator Foramen	0.900	<0.001	0.896	<0.001

Table 3. Intra-observer agreement of sex determination from individual morphological traits, scored on real and virtual specimens (Observer 1).

Agreement for assessments of sex (intra- and inter-observer as well as inter-modality) was examined using the Kappa statistic which determines consistency among observers (Landis and Koch 1977). Landis and Koch (1977: 165) describe the strength of agreement for the Kappa statistic as follows: 0.41 - 0.60 moderate agreement; 0.61 - 0.80 substantial agreement; 0.81 - 1.00 almost perfect agreement. This scale is used here for consistency, but clearly, values above 0.80 are desirable for any meaningful level of observer agreement.

Intra-observer agreement of traits scored between real and virtual specimens was substantial for Observer #1 (Kappa=0.692; p<0.001) and fair for Observer #2 (Kappa=0.322; p<0.001). Repeated scores by the same observer (#1) show substantial levels of agreement for real specimens (Kappa=0.667; p<0.001) and moderate agreement for virtual specimens (Kappa=0.573; p<0.001). Table 3 shows the results of intra-observer agreement on real and virtual specimens, for Observer #1, broken down by specific morphological trait.

Moderate levels of inter-observer agreement were found on morphological traits scored on the real and virtual specimens. Table 4 shows the results of inter-observer agreement, on real specimens and virtual specimens for each of the morphological traits.

Trait	Real Specimens		Virtual Specimens	
	Kappa	p	Kappa	p
Greater Sciatic Notch	0.634	<0.001	0.55	0.003
Ventral Arc	0.503	0.008	0.667	0.002
Subpubic Concavity	0.588	0.003	0.571	0.001
Medial Aspect of Ischiopubic Ramus	1	<0.001	0.311	0.053
Preauricular Sulcus	0.674	<0.001	0.76	<0.001
Acetabulum	0.5	0.005	0.449	0.004
Obturator Foramen	0.396	0.015	0.534	0.003

Table 4. Inter-observer agreement of sex determination from individual morphological traits, scored on real and virtual specimens.

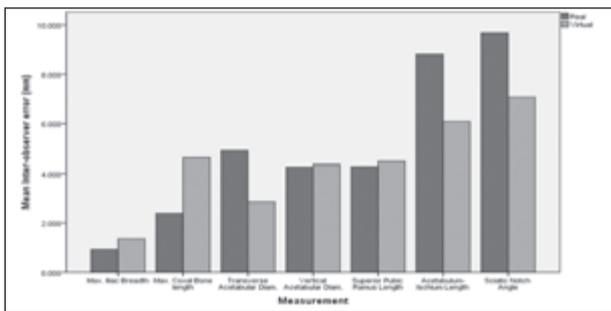


Figure 2. Percentage congruency (agreement) of sex determination between observers by morphological trait in real versus virtual specimens.

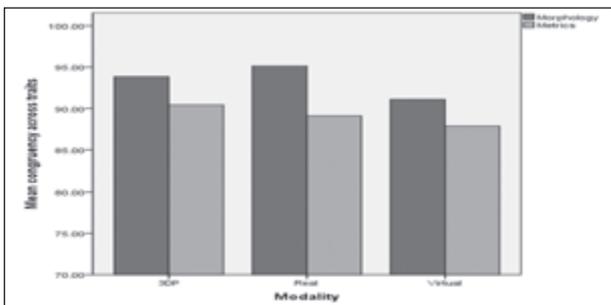


Figure 3. Average congruency of sex determination by modality (Observer 1).

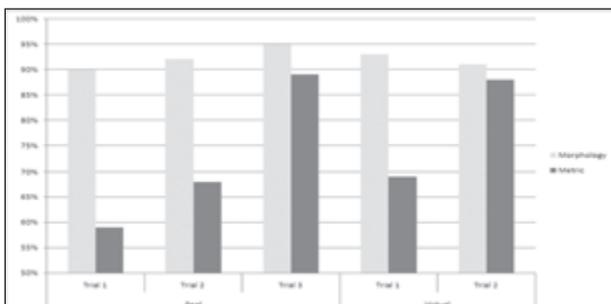


Figure 4. Overall congruency of sex determination by trial and modality (Observer 1).

No significant intra-observer error was demonstrated for the measurements ($t=0.827$; $df=289$; $p=0.409$) but there were significant differences between observers with the mean difference being 1.7mm on the real specimens and 2.2mm on the virtual specimens. Mean inter-observer error

comparison for the measurements for both the real and the virtual specimens is shown in figure 2. The magnitude of inter-observer error, however, is not significantly different between the real and the virtual specimens ($t=1.28$; $df=181$; $p=0.202$). That is to say, while there are observer differences in measurement, the modality does not seem to be a factor. The measures most susceptible to inter-observer error were the acetabulum-ischium length and the greater sciatic notch angle on real specimens, though the latter showed no significant difference between observers on the virtual specimens – likely resolved by easy angle measurement tools available in Rapidform.

Finally, an assessment of congruency of sex determination between each of the modalities was made. A high level of agreement was observed between overall sex determination on real versus virtual specimens (morphological traits: $Kappa=1.000$, $p<0.001$; metrics $Kappa=0.908$, $p<0.001$) and for real versus the 3DP models (morphological traits: $Kappa=1.000$, $p=0.008$; metrics $Kappa=0.696$, $p=0.053$) though the sample size was small for the latter.

The distribution of within case congruency (percentage of traits in agreement for sex determination) averaged across all trials is presented below for each of the modalities (Fig. 3). Also observed was a positive correlation in increased congruence with trials for the metrics although morphology is stable across trials (Fig. 4). This is interpreted as a learning curve for the

measurements with increased experience being a key factor and would explain in part, the mean error in measurements between observers.

Discussion and Conclusions

It is clear that some of the morphological features (e.g. preauricular sulcus) were difficult to discern on the virtual and 3DP models due to issues of precision from the scanning and smoothing in the 3D renders of those data. Metric assessments were more easily captured on the virtual model which made it possible to establish more consistent landmarks across multiple measurements from the same point. Despite this, the measurements taken on the virtual models showed less internal congruency across the sample, and greater variation when compared to the real specimens. This could be related to the level of experience of the observer or it could be a product of the smoothing effects of the scan data which may have compromised the specific identification of landmarks. Lastly, landmarks that are incorrectly placed on the virtual model, would affect all measurements that are taken from that particular point, thus error in individual measurements may not always be independent of one another in the 3D environment where landmark are 'snapped' to the surface of the 3D model.

Sholts et al. (2010) examined accuracy and precision of volume and surface measures of data from 3D laser scans of 5 crania, observing very small levels of both intra- and inter-observer error. In a second publication (Sholts et al. 2011) they compare the precision of landmark data on the 5 crania acquired from laser scans versus a Microscribe digitizer. While they concluded that both data sources were adequate for anthropological research purposes, they observed higher levels of precision using the digitizer, which they attributed to taphonomic and general preservation having a larger impact on the scan data and a researcher's ability to consistently identify appropriate landmark coordinates.

These results were consistent with Williams and Richtsmeier (2003) who similarly observed slightly better precision in landmarks taken on 22 mandibles when using digitizer versus CT data, and Park et al. (2006) who observed good results for craniometrics from laser scan data on 30 skulls. Park and colleagues (2006) did, however, observe that the measurements from the laser scan coordinates using the Polhemus stylus (essentially creating digitizer data) showed slightly smaller measurements than on the actual specimen. This was also observed by Allard (2006) who noted CT scan data produced slightly larger 3D models when compared to those produced by point cloud data collected using a Polhemus hand-held laser scanner. Most recently, Decker et al. (2011) observed a near perfect level of observer agreement in the assessment of sex from 3D rendered models of the pelvis from CT data; higher than the inter-observer agreement presented here.

The accuracy of sex determination techniques developed from the pelvis are well established. Methodologies focusing on single traits in the pelvis, while often resulting in high levels of accuracy, do not necessarily reflect the full sexual dimorphism of the pelvis and cannot be reliably applied to different populations. The development of methods founded on multiple relevant traits which are well defined and easy to recognize and to score, is therefore essential for osteological research (Bruzek 2002; Bruzek and Murail 2006; Ferembach et al. 1980). Bruzek and Murail caution that:

"observing traits or taking measurements of a single morphofunctional segment of the hip bone is not appropriate. The sexual dimorphism observed on a single segment is often influenced by size and thus, population-specific ... In fact, the variation in sexual dimorphism of one segment of the hip bone influences the variation in the other segments of the hip bone" (2006, 228).

The most accurate techniques will be those that consider all sexual dimorphism in the pelvis

by specifying a certain number of traits rather than by increasing the number of variables (Bruzek 2002; Bruzek and Murail 2006; Ferembach 1980). In the current study, those traits that seem most able to be reliably scored on real, virtual 3D and 3DP models include the greater sciatic notch, ventral arc and subpubic concavity.

The results from this study suggest that both morphological and metric-based sex determination using virtual and 3DP models of human pelvises are generally (but not always) congruent with those based on the real pelvis. While morphological assessments benefit from the use of the real skeletal material, the use of advanced imaging for metric assessments may be useful especially with regard to minimizing handling-induced damage to real bone. The differing levels of internal congruency on the virtual models highlight an area which requires further investigation. This could be indicative of a strength of having a 'real' (whether original or 3DP) element for handling. It may also be that the observer requires more extensive specialized training with the virtual system in advance of their observations. Only further research can identify the source of this discrepancy.

Increasingly, researchers are utilizing virtual and 3DP models in anthropological research. As the application of 'traditionally' developed techniques to virtual and 3DP models becomes more common, it is crucial to continue to test the congruence of techniques using these different modalities.

Acknowledgements

This research was supported by: University of Manitoba Graduate Fellowships; Social Sciences and Humanities Research Council of Canada; Canada Research Chairs program.

Bibliography

- Albanese, J. 2003. "A Metric Method for Sex Determination Using the Hipbone and the Femur." *Journal of Forensic Science* 48 (2):1-11.
- Allard, T. T. 2006. "The Role of 3D Printing in Biological Anthropology." MA Thesis, University of Manitoba.
- Allard, T. T., Sitchon, M., Sawatzky, R., and Hoppa, R. D. 2005. "Use of Hand-held Laser Scanning and 3D Printing for Creation of a Museum Exhibit." In *Proceedings of the 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage: Short and Project Papers*, edited by M. Mudge, N. Ryan, and R. Scopigno, 97-101.
- Arsuaga, J. L., and Carretero, J. M. 1994. "Multivariate Analysis of the Sexual Dimorphism of the Hip Bone in a Modern Human Population and in Early Hominids." *American Journal of Physical Anthropology* 93:245-57.
- Barker, T. M., Earwaker, W. J., and Lisle, D. A. 1994. "Accuracy of stereolithographic models of human anatomy." *Australas Radiol* 38:106-11.
- Bouyssié, J. F., Bouyssié, S., Sharrock, P., and Duran, D. 1997. "Stereolithographic models derived from x-ray computed tomography. Reproduction accuracy." *Surgical and Radiologic Anatomy* 19:193-9.
- Bruzek, J. 2002. "A method for visual determination of sex, using the human hip bone." *American Journal of Physical Anthropology* 117 (2):157-68.
- Bruzek, J., and Murail, P. 2006. "Methodology and reliability of sex determination from the skeleton." In *Forensic anthropology and medicine*, edited by A. Schmitt, E. Cunha, and J. Pinheiro, 225-42. Totowa: Humana Press.
- Buikstra, J. E., and Ubelaker, D. H. 1994. *Standards for Data Collection from Human Skeletal Remains*. Fayetteville: Arkansas Archaeological Survey Report.

Congruence of Methods for Determination of Sex using Real, Virtual and 3-D Printed Specimens
Julia Gamble, Amanda Blackburn and Robert D. Hoppa

- Choi, J. Y., Choi, J. H., Kim, N. K., Kim, Y., Lee, J. K., Kim, M. K., Lee, J. H., and Kim, M. J. 2002. "Analysis of errors in medical rapid prototyping models." *International Journal of Oral and Maxillofacial Surgery* 31:23-32.
- Decker, S. J., Davy-Jow, S. L., Ford, J. M., and Hilbelink, D. R. 2011. "Virtual Determination of Sex: Metric and Nonmetric Traits of the Adult Pelvis from 3D Computed Tomography Models." *Journal of Forensic Science*.
- Fantini, M., de Crescenzo, F., Persiani, F., Benazzi, S., and Gruppioni, G. 2008. "3D restitution, restoration and prototyping of a medieval damaged skull." *Rapid Prototyping Journal* 14 (5):318-24.
- Ferembach, D., Schwidetzky, I., and Stloukal, M. 1980. "Recommendations for age and sex diagnoses of skeletons." *Journal of Human Evolution* 9 (7):517-549.
- France, D. L. 1998. "Observational and Metric Analysis of Sex in the Skeleton." In *Forensic Osteology: Advances in the Identification of Human Remains*, edited by K. J. Reichs, 163-186. Springfield: Charles C. Thomas Publisher Ltd.
- Hjalgrim, H., Lynnerup, N., Liversage, M., and Rosenkint, A. 1995. "Stereolithography: potential applications in anthropological studies." *American Journal of Physical Anthropology* 97:329-33.
- Landis, J. R., and Koch, G. G. 1977. "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33 (1):159-74.
- Park, H.-K., Chung, J.-W., and Kho, H.-S. 2006. "Use of hand-held laser scanning in the assessment of craniometry." *Forensic Science International* 160 (2):200-06.
- Phenice, T. W. 1969. "A newly developed visual method of sexing in the os pubis." *American Journal of Physical Anthropology* 30:297-301.
- Pietrusewsky, M. 2008. "Metric Analysis of Skeletal Remains: Methods and Applications." In *Biological Anthropology of the Human Skeleton*, edited by M. A. Katzenberg and S. R. Saunders, 487-532. New York: John Wiley & Sons Inc.
- Richtsmeier, J. T., DeLeon, V. B., and Lele, S. 2002. "The promise of geometric morphometrics." *Yearbook of Physical Anthropology* 45:63-91.
- Rogers, T., and Saunders, S. 1994. "Accuracy of sex determination using morphological traits of the human pelvis." *Journal of Forensic Science* 39:1047-56.
- Schwartz, J. H. 1995. *Skeleton Keys: An introduction to human skeletal morphology, development, and analysis*. New York: Oxford University Press.
- Sholts, S. B., Wärmländer, S. K. T. S., Flores, L. M., Miller, K. W. P., and Walker, P. L. 2010. "Variation in the Measurement of Cranial Volume and Surface Area Using 3D Laser Scanning Technology." *Journal of Forensic Science* 55 (4):871-76.
- Sholts, S. B., Flores, L. M., Walker, P. L., and Wärmländer, S. K. T. S. 2011. "Comparison of Coordinate Measurement Precision of Different Landmark Types on Human Crania Using a 3D Laser Scanner and a 3D Digitiser: Implications for Applications of Digital Morphometrics." *International Journal of Osteoarchaeology* 21.
- Steyn, M., and Patriquin, M. L. 2009. "Osteometric sex determination from the pelvis – does population specificity matter?" *Forensic Science International* 191:113-115.
- Sutherland, L. D., and Suchey, J. M. 1991. "Use of the ventral arc in pubic sex determination." *Journal of Forensic Sciences* 26:501-11.
- Ubelaker, D. H., and Volk, C. G. 2002. "A test of the phenice method for the estimation of sex." *Journal of Forensic Science* 47:19-24.
- Walker, P. L. 2008. "Sexing Skulls: Using Discriminant Function Analysis of Visually Assessed

Traits.” *American Journal of Physical Anthropology* 136:39-50.

Walrath, D. E., Turner, P., and Bruzek, J. 2004. “Reliability test of the visual assessment of cranial traits for sex determination.” *American Journal of Physical Anthropology* 125 (2):132-37.

Weber, G. W. 2001. “Virtual anthropology (VA): A call for Glasnost in paleoanthropology.” *The Anatomical Record (Part B: New Anatomy)* 265:193-201.

Williams, F. L., and Richtsmeier, J. T. 2003. “Comparison of Mandibular Landmarks from Computed Tomography and 3D Digitizer Data.” *Clinical Anatomy* 16:494-500.