



The accuracy of 3D virtual bone models of the pelvis for morphological sex estimation

Kerri L. Colman¹ · Alie E. van der Merwe¹ · Kyra E. Stull^{2,3} · Johannes G. G. Dobbe⁴ · Geert J. Streekstra^{4,5} · Rick R. van Rijn⁵ · Roelof-Jan Oostra¹ · Hans H. de Boer^{6,7}

Received: 25 July 2018 / Accepted: 11 January 2019 / Published online: 24 January 2019
© The Author(s) 2019

Abstract

It is currently unknown whether morphological sex estimation traits are accurately portrayed on virtual bone models, and this hampers the use of virtual bone models as an alternative source of contemporary skeletal reference data. This study determines whether commonly used morphological sex estimation traits can be accurately scored on virtual 3D pelvic bone elements. Twenty-seven intact cadavers from the body donation program of the Amsterdam UMC, University of Amsterdam, were CT scanned; this data was used to produce virtual bone models. Thereafter, the dry bones were obtained. Three traits by Klales (2012) and five traits from the Workshop of European Anthropologists (WEA) (1980) were scored on the virtual bone models and their dry skeletal counterparts. Intra- and inter-observer agreement and the agreement between the scores for each virtual bone model-dry bone pair were calculated using weighted Cohen's kappa (K). For all Klales (2012) traits, intra- and inter-observer agreement was substantial to almost perfect for the virtual- and dry bones ($K = 0.62\text{--}0.90$). The agreement in scores in the virtual-dry bone pairs ranged from moderate to almost perfect ($K = 0.58\text{--}0.82$). For the WEA (1980) traits, intra-observer agreement was substantial to almost perfect ($K = 0.64\text{--}0.91$), but results were less unambiguous for inter-observer agreement ($K = 0.24\text{--}0.88$). Comparison of the scores between the virtual bone models and the dry bones yielded kappa values of $0.42\text{--}0.87$. On one hand, clinical CT data is a promising source for contemporary forensic anthropological reference data, but the interchangeability of forensic anthropological methods between virtual bone models and dry skeletal elements needs to be tested further.

Keywords Radiology · Virtual anthropology · Accuracy · Reliability · Sex estimation · Pelvis

✉ Kerri L. Colman
k.l.colman@amc.uva.nl

- ¹ Department of Medical Biology, Section Clinical Anatomy and Embryology, Amsterdam UMC, location AMC, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands
- ² Department of Anthropology, University of Nevada, Reno, Reno, Nevada, USA
- ³ Faculty of Health Sciences, Department of Anatomy, University of Pretoria, Pretoria, South Africa
- ⁴ Department of Biomedical Engineering and Physics, Amsterdam UMC, location AMC, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands
- ⁵ Department of Radiology, Amsterdam UMC, location AMC, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands
- ⁶ Department of Pathology, Amsterdam UMC, location AMC, University of Amsterdam, Meibergdreef 9, Amsterdam, The Netherlands
- ⁷ Department of Forensic Medicine, Netherlands Forensic Institute, The Hague, The Netherlands

Introduction

Sex estimation is a key component for creating a forensic anthropological biological profile as the estimation of the other biological profile elements (age at death, stature, and ancestry) relies heavily on the accurate estimation of sex. Mainly due to differences in reproductive function the pelvis is considered to be the most sexually dimorphic skeletal element in humans and therefore the most reliable skeletal element for sex estimation [1–7]. Both metric and morphological sex estimation methods exist, but the latter is often favored because of the ease of application and applicability across populations and time [8–10].

In order to be considered acceptable as evidence in a Court of Law, the accuracy of forensic anthropological sex estimation methods need to be known [11, 12]. For this, they should be derived from or tested on large and representative skeletal populations. Large known skeletal populations exist in numerous parts of the world (e.g. the USA, South Africa, Thailand) and have been used in development and validation studies [7,

13, 14]. However, in most European countries, no contemporary collections exist [15–20]. Specifically in the Netherlands, due to legislation and the high mean age of bodies donated to the body donation programs in medical schools [21, 22], it is impossible to compile a skeletal collection that is representative of the current population in the Netherlands.

Radiological data, such as computed tomography (CT) scans, derived from a clinical/hospital setting, might be used as a proxy since the number of scans generated on a daily basis guarantees a large, contemporary and representative source of data. Data acquired from a clinical radiological setting is actually a more accurate representation of the current population than most body donation programs since it is less affected by secular trends and age bias. Furthermore, reconstructions that are based on the CT scan data can provide a precise three-dimensional (3D) model that represents the original skeletal element [23]. However, in order for CT scans to be a feasible alternative to traditional skeletal collections, the morphological traits indicating sex also need to be accurately portrayed in the 3D reconstructions and their associated sex estimation scores need to correspond to the scores recorded on the dry skeletal element.

Some studies have already explored whether morphological methods can be applied on radiological data sources, however, the traits investigated were primarily scored on a binary scale (i.e. absent/present scores) [24, 25]. It therefore remains uncertain if the often-subtle morphological features defined on ordinal scales (scales of 1 to 5 or –2 to 2) can be scored accurately on virtual bone models. Consequently, one cannot readily assume that morphological sex estimation methods that are traditionally applied to dry skeletal remains perform equally when applied to virtual 3D skeletal elements.

The current study aims to determine whether commonly used morphological sex estimation methods that utilize ordinal scales can be accurately scored on reconstructed virtual 3D pelvic bone elements.

Material and methods

Twenty-seven (13 males, 14 females) fully intact cadavers were randomly selected from the body donation program of the Amsterdam Universitair Medische Centra (UMC), University of Amsterdam, Department of Medical Biology, section Clinical Anatomy and Embryology. The mean age at death was 78 years (range 52–94 years) for males and 80 years (range 64–100 years) for females (Fig. 1).

All cadavers were CT scanned while fully fleshed using a standard patient scanning protocol (120 kV, 150 mAs, slice thickness 0.9 mm, increment 0.45 mm, reconstruction

kernel D (bone filter)) on a Philips Brilliance 64 (Philips Medical Systems, Best, The Netherlands). The scan data were used to segment 3D virtual bone models of the os coxae using dedicated in-house research software. Specifics regarding this software package and segmentation process can be found in Dobbe (2011) [26] and Colman (2017) [23]. Following CT scanning, the bodies were macerated to obtain the dry os coxae by using a straightforward, yet effective, method of removing the majority of soft tissue and then submerging the os coxae into boiling water.

Morphological sex estimation traits, namely the ventral arc (VA), subpubic concavity (SPC) and the medial aspect of the ischial pubic ramus (MA), being the three Phenice traits expanded to an ordinal scale as described by Klales (2012) [8], and the pre-auricular sulcus (PAS), greater sciatic notch (GSN), pubic angle (PA), arc compose (AC), and the ischial body (IB) as described by the Workshop of European Anthropologists (WEA) (1980) [27, 28] were selected for investigation. These five WEA (1980) traits were selected due to the range of complexity that they provide. In addition, the selected traits are amongst the ones most commonly used for sex estimation in the anthropological community [13, 29–31]. All selected traits, both Klales (2012) and the specific WEA (1980) traits, were scored on both the virtual bone models and the dry bones. Scoring was performed using a five-tier hyper-feminine to hyper-masculine scale ranging between –2 and 2 for the WEA (1980) characteristics and 1 to 5 for the features as described by Klales (2012). See WEA (1980) and or Klales (2012) for detailed descriptions associated with each score. The virtual bone models were scored using the aforementioned dedicated in-house research software, as well as MeshLab [32], an open source online commercial viewing software, to determine whether differences in software packages would result in different scores or accuracies. Scoring took place twice by two independent observers with a minimum of 1-week lapse between repeat observations.

Intra- and inter-observer agreement for scores on the virtual bone models and the dry bone counterparts (i.e., reliability), as well as the agreement between the scores for each virtual bone model-dry bone pair (i.e. accuracy), were calculated using Cohen's (1968) weighted kappa (K) [33]. Levels of agreement were interpreted according to the thresholds defined by Landis and Koch (1977) [34]: $K < 0$ indicates less than chance agreement, $K = 0.01$ to 0.20 indicates slight agreement, $K = 0.21$ to 0.40 indicates fair agreement, $K = 0.41$ to 0.60 indicates moderate agreement, $K = 0.61$ to 0.80 indicates substantial agreement and $K = 0.81$ to 1.0 indicates almost perfect to perfect agreement. All statistical analyses were done using R, version 3.3.0.

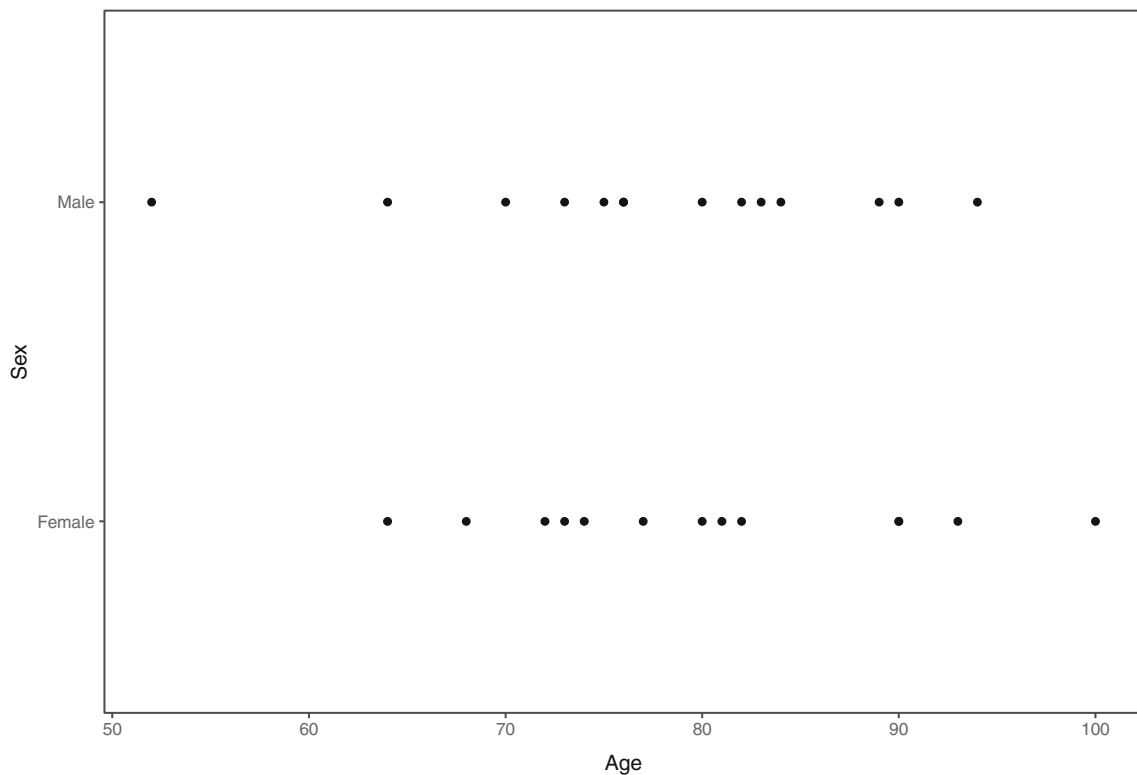


Fig. 1 Demographic profile of the 27 (13 males, 14 females) fully intact cadavers included in this study

Results

Sex estimation features based on the Klales (2012) traits

Intra- and inter-observer agreement

Cohen’s weighted kappa (K) values indicate substantial to almost perfect agreement for the intra- and inter-observer agreement for the three Klales (2012) traits when scored in isolation on the dry bones (K = 0.78–0.89 and 0.61–0.77), as well as on the virtual bone models using both software packages (in-house software K = 0.78–0.84 and 0.62–0.72; MeshLab K = 0.84–0.90 and 0.62–0.80). See Table 1 for the scores per trait and modality.

Agreement between the virtual bone models and their dry bone counterparts

The agreement between the Klales (2012) traits scored on the virtual bone models and their dry bone counterparts indicate substantial to almost perfect agreement for all features (K = 0.74–0.82) when using the in-house research software. The level of agreement was slightly less with moderate to substantial agreement (K = 0.58–0.76) when using MeshLab (see Table 2).

Sex estimation features based on the WEA (1980) traits

Intra- and inter-observer agreement

The intra-observer agreement for the dry bone elements for all five WEA (1980) traits was substantial to almost perfect (K = 0.76–0.88). Similar values were found on the virtual bone models, using the in-house research software (K = 0.64–0.91

Table 1 Cohen’s weighted kappa (K) values for the levels of intra- and inter-observer agreement for Klales (2012) traits when scored on dry bones and virtual bone models

	VA	SPC	MA
Kappa values for intra-observer agreement			
Dry bone	0.89	0.81	0.78
Virtual bone in-house software	0.84	0.84	0.78
Virtual bone MeshLab	0.90	0.86	0.84
Kappa values for inter-observer agreement			
Dry bone	0.67	0.77	0.61
Virtual bone in-house software	0.72	0.69	0.62
Virtual bone MeshLab	0.80	0.68	0.62

VA ventral arc, SPC sub-pubic concavity, MA medial aspect ischial-pubic ramus

Table 2 Cohen's weighted kappa (K) values indicating levels of agreement for Kales (2012) traits between virtual bone models and their dry bone counterparts

	VA	SPC	MA
	Kappa values virtual bone model vs. dry bone		
Virtual in-house software vs. dry bone	0.74	0.84	0.82
Virtual MeshLab vs. dry bone	0.58	0.76	0.70

VA ventral arc, SPC sub-pubic concavity, MA medial aspect ischial-pubic ramus

for all five traits). Using MeshLab, the scores of all but one trait (the IB) indicated almost perfect agreement ($K = 0.82$ – 0.86). The intra-observer agreement for the ischial body was “fair” ($K = 0.40$).

Results were less unambiguous for the inter-observer agreement. In the dry bone elements, Cohen's weighted kappa values indicated substantial to almost perfect agreement for PAS, PA, and AC ($K = 0.70$ – 0.88). Inter-observer agreement for the GSN was moderate ($K = 0.51$) and for the IB fair ($K = 0.24$). In the virtual models (using both software packages), substantial to almost perfect inter-observer agreement was found for the PAS, GSN, and PA (in-house software $K = 0.65$ – 0.83 , MeshLab $K = 0.68$ – 0.83), while the scores for AC indicated moderate agreement ($K = 0.54$ and $K = 0.55$). The IB performed worst with kappa values of 0.33 and 0.05 for the in-house software and MeshLab, respectively (see Table 3).

Agreement between the virtual models and their dry bone counterparts

The agreement between the virtual bone models and their dry bone counterparts for the five WEA (1980) traits indicate substantial to almost perfect agreement ($K = 0.68$ – 0.85) for three of the five traits, namely the GSN, PA and AC when using the in-house research software (see Table 4). PAS and IB showed

Table 3 Cohen's weighted kappa (K) values indicating levels of intra- and inter-observer agreement for WEA (1980) traits when scored on dry bones and virtual bone models

	PAS	GSN	PA	AC	IB
Kappa values for intra-observer agreement					
Dry bone	0.93	0.76	0.80	0.88	0.80
Virtual in-house software	0.71	0.91	0.80	0.81	0.64
Virtual MeshLab	0.86	0.82	0.84	0.83	0.40
Kappa values for inter-observer agreement					
Dry bone	0.88	0.51	0.76	0.70	0.24
Virtual in-house software	0.65	0.83	0.82	0.55	0.33
Virtual MeshLab	0.80	0.68	0.83	0.54	0.05

PAS pre-auricular sulcus, GSN greater sciatic notch, PA pubic angle, AC arc compose, IB ischial body

Table 4 Cohen's weighted kappa (K) values indicating levels of agreement for WEA (1980) traits when scored on virtual bone models and their dry bone counterparts

	PAS	GSN	PA	AC	IB
	Kappa values virtual bone model vs. dry bone				
Virtual in-house software vs. dry bone	0.42	0.80	0.85	0.68	0.50
Virtual MeshLab vs. dry bone	0.81	0.74	0.87	0.80	0.33

PAS pre-auricular sulcus, GSN greater sciatic notch, PA pubic angle, AC arc compose, IB ischial body

moderate agreement ($K = 0.42$ – 0.50). Using MeshLab, the scores of all but one trait (IB) indicated substantial to almost perfect agreement ($K = 0.74$ – 0.87). IB showed fair agreement ($K = 0.33$).

An example of the differences/similarities observed when comparing the dry bone- and virtual 3D models is demonstrated in Fig. 2, exemplifying a trait that scored well (pubic angle), and Fig. 3, illustrating the differences observed in a feature that performed poorly (pre-auricular sulcus).

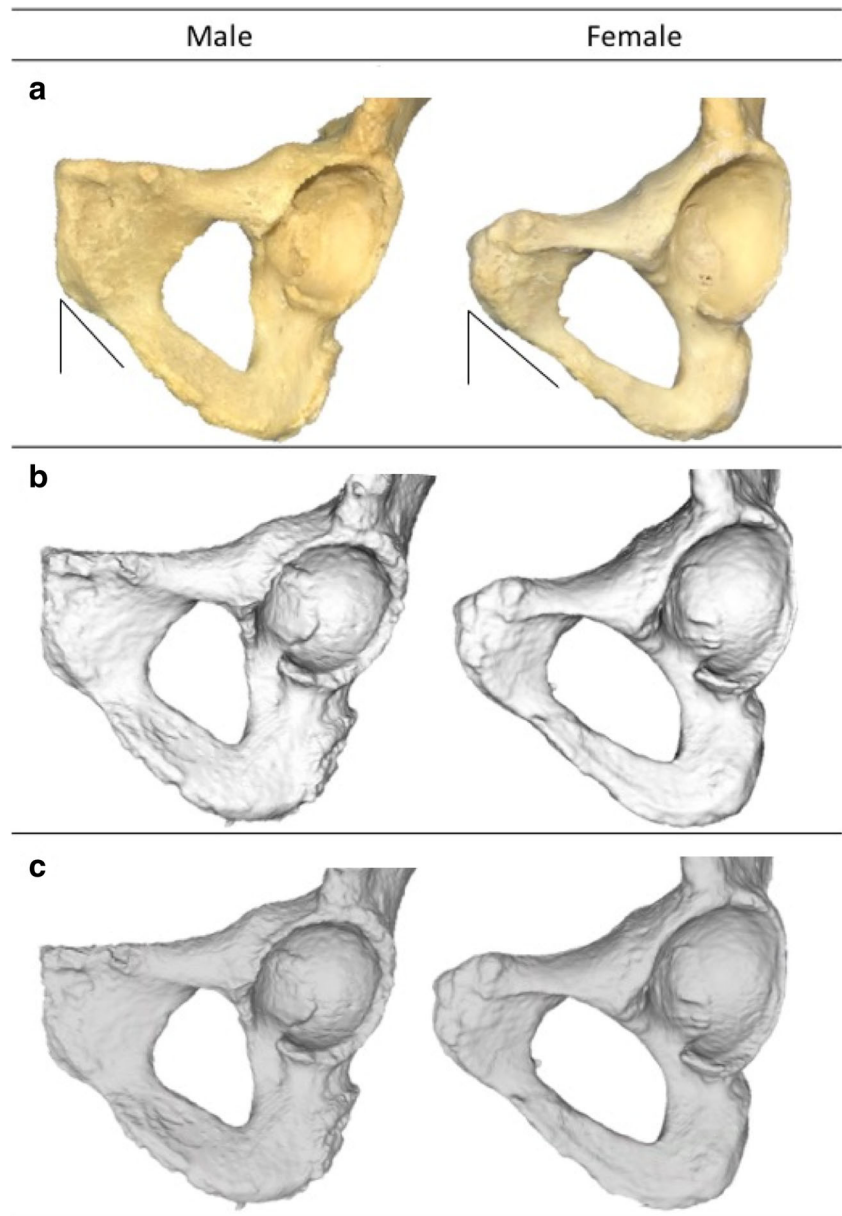
Discussion

Intra- and inter-observer agreement results indicate that all three Kales (2012) traits can be scored reliably on both the virtual bone models (regardless of which viewer software package is used) and on the dry bone counterparts. Notable is that the intra- and inter-observer agreement found in this study were comparable to those in the original Kales (2012) publication [8]. Furthermore, the results from this study are in line with previous research conducted by Johnstone-Belford (2018) [24], which found that the intra-observer variation is limited when the binary Phenice traits [35] are scored on virtual bone models. Additionally, scores assigned for all three Kales (2012) traits on the virtual bone models showed acceptable accuracy scores. VA shows slightly reduced accuracy when viewed in MeshLab, but the kappa value of 0.58 is still considered acceptable by the authors. Based on these results, the authors conclude that the Kales (2012) morphological sex estimation technique can be used interchangeably between virtual bone models and dry skeletal elements. Consequently, classification accuracies of the published method should be valid on virtual bone models and additionally, virtual data from clinical CT scans can be used to determine population specific accuracies.

The same does not hold for all selected traits incorporated in the WEA (1980) morphological sex estimation technique.

The IB scored poorly overall and proved to be a difficult trait to score on both the virtual bone models and the dry bones. In this study, the IB was found to be neither reliable (i.e. lack of intra- and inter-observer agreement) or accurate

Fig. 2 A male and female example of a trait that performed reliably is the pubic angle on dry bone (**a**) and on two virtual 3D models; the in-house research software package (**b**) and MeshLab (**c**). The pubic angle is indicated by the lines in **a**



(i.e. lack of agreement between the scores in the virtual bone model-dry bone pair); thus, the authors feel that this should prompt a reconsideration of the IB as a sex estimation trait.

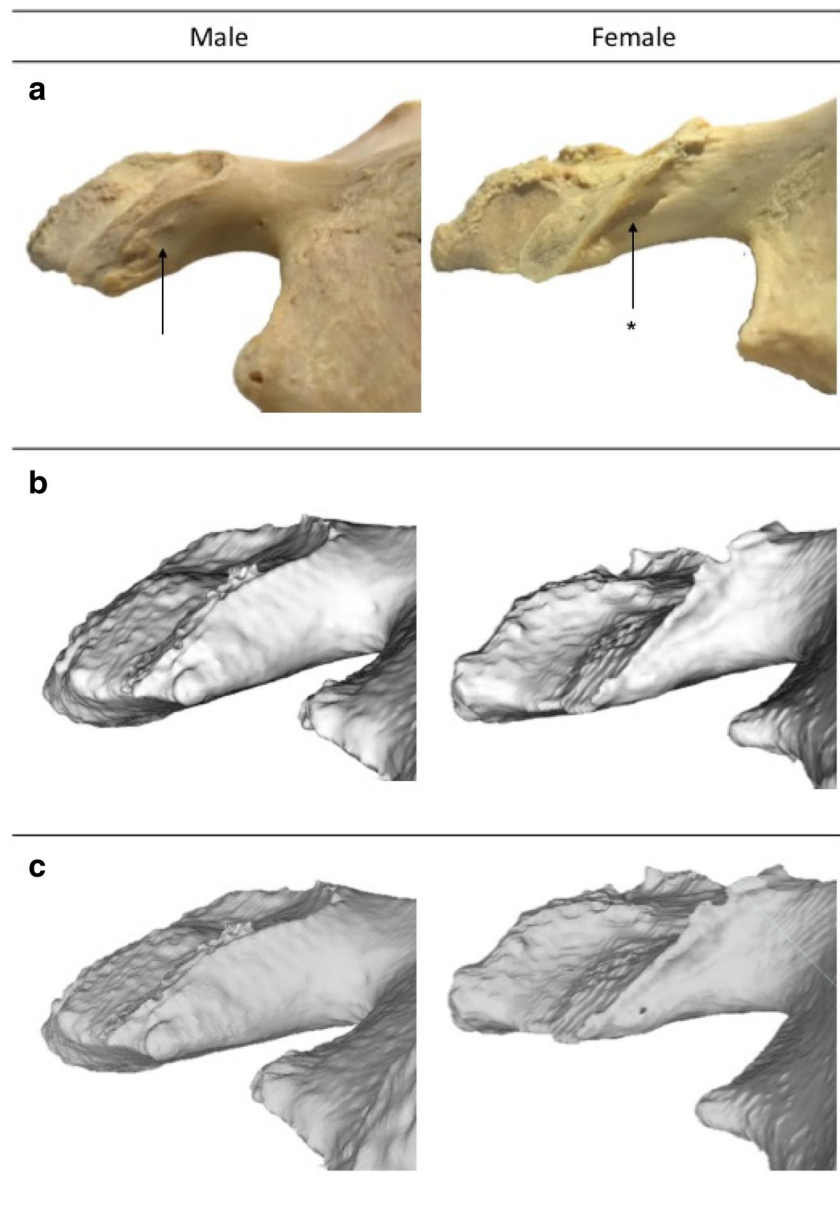
Intra-observer agreement for the four remaining WEA traits was substantial for both the virtual bone models (regardless of which viewer software package is used) and dry bone counterparts. However, inter-observer agreement scores showed that only the PAS and the PA show acceptable inter-observer agreement on the dry bone counterparts and on both the virtual bone models (regardless of which viewer software package is used). Inter-observer agreement levels were not acceptable for the GSN on dry bone and for the AC on virtual bone models (using both viewer software packages). Due to the lack of literature in which the intra- and inter-observer errors for each WEA (1980) trait is reported individually, the

authors were unable to compare the presented results to previous reliability performances of the separate traits.

Based on the agreement of scores between the virtual bone model-dry bone pairs, the sex estimation scores on virtual bone models for the GSN, PA and AC accurately correspond to those on the dry skeletal element (regardless of which viewer software package is used). However, only the PA can be used interchangeably between virtual models and dry skeletal elements, since the GSN and AC are disqualified on the basis of their lack of inter-observer agreement.

Although the PAS performed reliably (i.e. intra- and inter-observer agreement), conflicting accuracy (i.e. virtual bone model-dry bone pair agreement between scores) results were found when using different viewing software packages. Apparently, the use of a specific viewer software package

Fig. 3 A male and female example of a trait that performed poorly is the pre-auricular sulcus, on dry bone (**a**) and on virtual 3D models; the in-house research software package (**b**) and MeshLab (**c**). The arrows indicate the region of interest with the sulcus being present (*) on the female pelvis. Note that the pre-auricular sulcus is clearly visible on the dry bone, but invisible when viewed with the in-house software and poorly defined when looking at the reconstruction with MeshLab



has a substantial effect on the visualization of this trait (Fig. 2). A potential explanation might be that the scoring of the PAS on dry bones traditionally involves palpating the region of interest and using a slanted light source to view the depth of the sulcus. The in-house research software does not provide an artificial light source to visualize the sulcus, thereby hampering the sex estimation. In contrast, the agreement scores increased when using MeshLab, which does provide a light source.

For the WEA (1980) traits, it can be concluded that only the PA can be scored reliably and accurately on virtual bone models, regardless of the viewing software. If MeshLab is used as a viewing software, then PAS also performs reliably and accurately. This means that the classification accuracies of these traits can be applied to virtual bone models and that

clinical CT scans can be used as a source for the development of their population-specific accuracies.

In this study, a single segmentation method was used, which may be considered a limitation since a different segmentation method may result in slightly different virtual bone models [36]. Since these differences are generally small, it is unlikely that this would have an effect on scoring morphological traits. It is important to note that this study focusses on the reliability and accuracy of the sex estimation traits and not on the reliability and accuracy of sex estimation itself. The latter would require a much larger study population with a less skewed age distribution. The skewed age distribution might initially be viewed as a limitation of the study, however, this is nullified by the aim of the current study. As a matter of fact, the relatively old age of the used population presents

somewhat of a “worst case scenario” since the aging effect on joint surfaces and areas of muscle attachments might interfere with the segmentation process and consequently the anthropological analysis. It is therefore expected that the studied traits will perform similarly or better in younger individuals. Another limitation may be the use of only five WEA traits, however as previously mentioned, these five were chosen specifically for the range of complexity that they provide.

This study assumes that the maceration process did not influence the morphology of the scored Klales (2012) and WEA (1980) traits. This is based on the absence of any published data on a potential change in these traits due to the maceration process. Also, since the dry bone elements were taken as the golden standard to which the virtual bone models were compared, this potential change in shape does not influence the results of this study.

The findings in this study further support the recently initiated and progressing shift towards virtual forensic anthropology, but also reveals limitations regarding the applicability of certain traits to virtual bone models. To our knowledge, this is the first study that investigates the accuracy of individual morphological traits on virtual bone models in comparison to their dry bone counterparts. Previous studies have investigated the accuracy of virtual bones; however, these studies did so by using metric sex estimation methods [30, 37, 38].

The authors believe this study to be a useful addition to current literature, as it illustrates the feasibility of using virtual bone models, derived from clinical CT scans, as a proxy for reference populations, while simultaneously showing that caution is advised when interchangeably applying traditional forensic anthropological methodology on virtual bone models. As demonstrated in this study, not all traits within a given technique achieve the same level of agreement. Thus, it is essential to study the agreement between the traits of interest before applying morphological sex estimation techniques interchangeably on virtual bone models and the dry skeletal elements.

Conclusion

The Klales (2012) morphological sex estimation technique is as reliable on virtual bone models (regardless of which viewer software package is used) as it is on dry pelvic bones. Furthermore, the sex estimation scores on the virtual bone models agree sufficiently with those on dry bone. The same does not hold for the five investigated WEA (1980) morphological sex estimation traits which show almost invariably limited reliability and/or limited accuracy. Only the PA (and the PAS when viewed in MeshLab) showed acceptable results in terms of reliability and accuracy.

This study shows that clinical CT data can be used as a source for contemporary and population-specific forensic

anthropological data, but that the agreement between virtual bone models and the dry skeletal elements should always be considered before forensic anthropological morphological methods are applied interchangeably.

Acknowledgments The authors would like to thank Aubrey van het Reve and Hannah Crijns for their participation in this study.

Compliance with ethical standards

Conflict of interest J.G.G. Dobbe is a developer of the in-house segmentation software that was used in this study. The other authors declare no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Phenice IW (1969) A newly developed visual method of sexing the os pubis. *Am J Phys Anthropol* 30:183–186
2. Krogman W, İşcan M (1986) The human skeleton in forensic science. CC Thomas, Springfield
3. MacLaughlin S, Bruce M (1986) Population variation in sexual dimorphism in the human innominate. *Hum Evol* 1:221–231
4. Rogers T, Saunders S (1994) Accuracy of sex determination using morphological traits of the human pelvis. *J Forensic Sci* 39:1047–1056
5. Ubelaker DH, Volk CG (2002) A test of the phenice method for the estimation of sex. *J Forensic Sci* 47:19–24
6. Bruzek J (2002) A method for visual determination of sex, using the human hip bone. *Am J Phys Anthropol* 117:157–168
7. Walker PL (2005) Greater sciatic notch morphology: sex, age, and population differences. *Am J Phys Anthropol* 127:385–391. <https://doi.org/10.1002/ajpa.10422>
8. Klales AR, Ousley SD, Vollner JM (2012) A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. *Am J Phys Anthropol* 149:104–114. <https://doi.org/10.1002/ajpa.22102>
9. Klales AR (2016) Secular change in morphological pelvic traits used for sex estimation. *J Forensic Sci* 61:295–301. <https://doi.org/10.1111/1556-4029.13008>
10. Kenyhercz MW, Klales AR, Stull KE, McCormick KA, Cole SJ (2017) Worldwide population variation in pelvic sexual dimorphism: a validation and recalibration of the Klales et al. method. *Forensic Sci Int* 277:e1-259.e8
11. (1993) *Daubert v. Merrel Dow Pharmaceuticals, Inc.* 113 SCt 2786. U.S. LEXIS 4408
12. Grivas CR, Komar DA (2008) Kumho, Daubert, and the nature of scientific inquiry: implications for forensic anthropology. *J Forensic Sci* 53:771–776. <https://doi.org/10.1111/j.1556-4029.2008.00771.x>

13. Patriquin ML, Loth SR, Steyn M (2003) Sexually dimorphic pelvic morphology in South African whites and blacks. *Homo* 53:255–262. <https://doi.org/10.1078/0018-442x-00049>
14. King CA (1997) Osteometric assessment of 20th century skeletons from Thailand and Hong Kong. Universal-Publishers
15. Maat GJ, Maes A, Aarents MJ, Nagelkerke NJ (2006) Histological age prediction from the femur in a contemporary Dutch sample. The decrease of nonremodeled bone in the anterior cortex. *J Forensic Sci* 51:230–237. <https://doi.org/10.1111/j.1556-4029.2006.00062.x>
16. Hens SM, Rastelli E, Belcastro G (2008) Age estimation from the human os coxa: a test on a documented Italian collection*. *J Forensic Sci* 53:1040–1043. <https://doi.org/10.1111/j.1556-4029.2008.00818.x>
17. Eliopoulos C, Lagia A, Manolis S (2007) A modern, documented human skeletal collection from Greece. *Homo* 58:221–228. <https://doi.org/10.1016/j.jchb.2006.10.003>
18. Gapert R, Black S, Last J (2009) Sex determination from the occipital condyle: discriminant function analysis in an eighteenth and nineteenth century British sample. *Am J Phys Anthropol* 138:384–394. <https://doi.org/10.1002/ajpa.20946>
19. Cardoso HF (2006) Brief communication: the collection of identified human skeletons housed at the Bocage Museum (National Museum of Natural History), Lisbon, Portugal. *Am J Phys Anthropol* 129:173–176. <https://doi.org/10.1002/ajpa.20228>
20. Maat GJR (2002) Citizens buried in the ‘Sint Janskerkhof’ of the ‘Sint Jans’ cathedral of ‘s-Hertogenbosch in The Netherlands, ca. 1450 and 1830-1858 AD. In: *Anthropologica BS* (ed) Leiden, pp 1–62
21. Bolt S, Venbrux E, Eisinga R, Kuks JB, Veening JG, Gerrits PO (2010) Motivation for body donation to science: more than an altruistic act. *Ann Anat* 192:70–74
22. Bolt S, Eisinga R, Altena M, Venbrux E, Gerrits PO (2013) Over my dead body: body donation and the rise in donor registrations in the Netherlands. *OMEGA J Death Dying* 66:57–77
23. Colman KL, Dobbe JG, Stull KE et al (2017) The geometrical precision of virtual bone models derived from clinical computed tomography data for forensic anthropology. *Int J Legal Med* 131:1155–1163. <https://doi.org/10.1007/s00414-017-1548-z>
24. Johnstone-Belford E, Flavel A, Franklin D (2018) Morphoscopic observations in clinical pelvic MDCT scans: assessing the accuracy of the Phenice traits for sex estimation in a Western Australian population. *J Forensic Radiol Imaging* 12:5–10
25. Decker SJ, Davy-Jow SL, Ford JM, Hilbelink DR (2011) Virtual determination of sex: metric and nonmetric traits of the adult pelvis from 3D computed tomography models. *J Forensic Sci* 56:1107–1114. <https://doi.org/10.1111/j.1556-4029.2011.01803.x>
26. Dobbe JG, Strackee SD, Schreurs AW et al (2011) Computer-assisted planning and navigation for corrective distal radius osteotomy, based on pre- and intraoperative imaging. *IEEE Trans Biomed Eng* 58:182–190. <https://doi.org/10.1109/TBME.2010.2084576>
27. Workshop of European Anthropologists (1980) Recommendations for age and sex diagnoses of skeletons. *J Hum Evol* 9:517–549. [https://doi.org/10.1016/0047-2484\(80\)90061-5](https://doi.org/10.1016/0047-2484(80)90061-5)
28. Maat GJR, Panhuysen RGAM, Mastwijk RW (1999) Manual for the physical anthropology report. Barge’s *Anthropologica Leiden*
29. Stull KE, Kenyhercz MW, L’Abbe EN (2013) Non-metric cranial and pelvic traits as a measure of sexual dimorphism in a modern South African population. *Am J Phys Anthropol* 150:266–301. <https://doi.org/10.1002/ajpa.22247>
30. Chapman T, Lefevre P, Semal P et al (2014) Sex determination using the Probabilistic Sex Diagnosis (DSP: Diagnose Sexuelle Probabiliste) tool in a virtual environment. *Forensic Sci Int* 234:189. e1-8
31. Ubelaker D, Buikstra JJAASR (1994) Standards for data collection from human skeletal remains. 44: 206
32. Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G. (2008) Meshlab: an open-source mesh processing tool. *Eurographics Italian chapter conference*. pp 129–36
33. Cohen J (1968) Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychol Bull* 70:213–220
34. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *biometrics* 33:159–174
35. Phenice TW (1969) A newly developed visual method of sexing the os pubis. *Am J Phys Anthropol* 30:297–301. <https://doi.org/10.1002/ajpa.1330300214>
36. van Eijnatten M, van Dijk R, Dobbe J et al (2017) CT image segmentation methods for bone used in medical additive manufacturing
37. Franklin D, Cardini A, Flavel A, Kuliukas A, Marks MK, Hart R, Oxnard C, O’Higgins P (2013) Concordance of traditional osteometric and volume-rendered MSCT interlandmark cranial measurements. *Int J Legal Med* 127:505–520. <https://doi.org/10.1007/s00414-012-0772-9>
38. Stull KE, Tise ML, Ali Z, Fowler DR (2014) Accuracy and reliability of measurements obtained from computed tomography 3D volume rendered images. *Forensic Sci Int* 238:133–140. <https://doi.org/10.1016/j.forsciint.2014.03.005>