4-2023

Error Rates in Sex Determination of Human Skeletal Remains

Vanessa Slone

Error Rates in Sex Determination of Human Skeletal Remains

An honors thesis presented to the
Department of Anthropology,
University at Albany, State University of New York
In partial fulfillment of the requirements
for graduation with Honors in Human Biology
and
graduation from the Honors College

Vanessa Slone

Research Advisor: John Polk, Ph. D

April 2023

Abstract

The possible fallibility of forensic methods has been under scrutiny for over a decade. Left out of the initial reviews were methods within forensic anthropology. A literature review was conducted to examine modern methods in determining the biological sex of human skeleton remains and their associated error rates in making a correct determination. Results showed no significant correlation between the number of traits being considered the subsequent error rate in determining sex. The skull was shown to be the least accurate in determining biological sex with the pelvis showing the lowest rate of error. Finally, between regression line-based methods and observer-based methods, observer-based methods had a lower rate of error.

Acknowledgments

I have many people to thank for their tireless effort and support. First and foremost, I would like to thank Dr. John Polk for his constant assistance, both in helping me outline my thesis question, as well as assisting in all stages of data gathering and writing. Without his help, I see no way that I could have completed this thesis. I would also like to thank Sarah Smith, who worked beside me and answered my many questions.

I'd also like to thank Susannah Hall, Sonia Dobronravov, and Sarah Smith again, for being unparalleled friends and supporters throughout my years at SUNY Albany. I could not have gotten where I am without them.

Finally, I'd like to thank my family, who have done nothing but support and encourage me to do all that I strive for, with no plan to stop anytime soon.

Table of Contents

Introduction

According to a Bureau of Justice Statistics Special Report conducted in 2004, approximately 4,400 human remains are recovered every year. And, of those 4,400, approximately 1,000 are unidentified (1). If human remains are discovered in stages of severe decomposition, or if DNA sequencing is unable to be conducted due to compromised or missing tissue, then forensic anthropology may be employed to identify the remains. Identification of human remains using forensic anthropology employs the matching of ante-mortem and post-mortem characteristics. A profile of the skeletal traits is created to make comparisons and attempt to identify the deceased. A major aspect of such a profile is the biological sex of the skeletal remains.

Forensic anthropology is a method that relies heavily on the experience of the analyst, and therefore, is open to the possibility of error, as are all sciences. Validation studies of common forensic anthropological methods are widespread in the literature.

All scientific fields should be vigilant about validating their methods, even long-established methods. Forensic science is no different, and, because of the adjacent nature of forensic science with the justice system, it is even more of a pressing question. No current method of identifying human remains is free of error, and it is always worth remembering that these errors have long-term consequences, especially in a court setting.

The 2009 National Academy of Sciences report on forensic science kickstarted an examination of long-standing practices in forensic science (2). The report found that many forensic subfields did not have sufficient scientific evidence backing up their admissibility in United States courts (3). Largely left out of this groundbreaking report, were methods in forensic anthropology. Anthropologists have taken it upon themselves to vet the technique of their field.

Previous research has been conducted on intra- and interobserver error rates in osteometric measurements (4). Other research has examined methods in age estimation of human remains (5), as well as height estimation (6). However, a comprehensive examination into the errors of determining biological sex is overdue.

This paper aims to examine previous studies in errors relating to determination of biological sex of human remains. From previous studies, general trends in sex determination errors will be identified. These trends may provide insight onto the efficacy of certain aspects in determining biological sex and give guidance as to which techniques may be most effective in the future, and which techniques may need re-examination.

Methods

A literature review was conducted, focusing on papers that found a calculated error rate in determining the sex of skeletal remains. Papers were limited to those published after 2000 as they would reflect the most up-to-date methods used by forensic anthropologists, and the methods that would likely be used by an expert witness in court. Sources were also limited to those that used skeletons from the modern era, as those would most closely simulate conditions of remains in a modern forensic setting.

For errors in sex determination the following data was collected from each paper; sample size, error rate, year of publication, portion of skeleton being reviewed, number of traits being considered, whether metric or non-metric methods were being used, and, finally, notes on the population under review.

Many previous researchers have emphasized the importance of making observations about remains in the context of their specific population (7-13). Note, the population of the skeletal remains did not factor into calculations, as an expert witness in a U.S. court would not have the population context that participants in these studies did. However, they were noted for the sake of data transparency and for possible future calculations. While populations were noted, analysis was not conducted on error rates and populations. This is because many populations were only represented in one study, and therefore, any possible biases or errors in the singular study could have affected population specific analysis. Studies examining the complex interactions of ethnicity and/or race with osteometric data are well represented in the literature (7-13)

An average error rate was calculated from all error rates for different sections of the skeleton. Pearson's scores were done between the number of traits considered and error rates and a correlation matrix was also conducted using Jamovi 2.3.21. Finally, comparative tests were done for non-metric vs metric methods and their subsequent error rates.

Results

All gathered data was compiled into a master table in order to better see trends in areas of interest (Table 1). Areas of interest included the portion of the skeleton under review in each study, the number of traits considered, the methodology of determining biological sex, and the error rate of determining biological sex. Other aspects of each study were also noted for the sake of transparency of data; these include a citation to the original study, the sample size of the study, the year the study was conducted, and a note about the population makeup of each study. From the master table, variables were graphed to better visualize possible trends between differing independent variables and calculated error rates in determining biological sex.

Several general trends were identified as well as an absence of significant correlations. First examined, and of the highest interest, was the possible correlation between the number of traits being considered and the calculated error rate of determining biological sex. The number of traits under consideration were plotted along with the error rate from the associated study (Fig.1). The Pearson's Correlation Coefficient (r^2) between these variables was 0.0762. This score indicates that only 7% of the dependent data (error rate) could be explained by the chosen independent variable (number of traits). There is evidence of minimal linkage between the two. Therefore, it was necessary to run a correlation matrix (Table. 2). The p-score that resulted from the correlation matrix was 0.075, which can be rejected according to the null hypothesis. According to the gathered data, there is no significant correlation between the number of traits under review and the error rate in determining biological sex.

Another question under review was the portion of the skeleton and how that compared the average error rate in determining biological sex. A bar graph showing the portion of the skeleton and the average overall error rate was created (Fig 2). Fig. 2 shows that the skull had the highest average error rate at 19.04%. The skull is followed by the clavicle (13.75%), long bones (12.34%), scapulae (11.93%), and finally the pelvis at 7.95%.

From the general areas of interest examined in Fig. 2, a more detailed examination of each skeletal portion was conducted. Fig. 3 is a bar graph examining the average error rate based on the section of the skull under review. Considering the skull as a whole proved to be the most accurate based on gathered data, with an average error rate of 15.15%. Examining the mandible exclusively had an average error rate of 22.75%, and examining the cranium exclusively had an average error rate of 23.30%.

Next, the long bones were examined in greater detail. Fig. 4 is a bar graph showing the long bones and the average error rate associated with each bone. The ulna had the highest average error rate at 14.87%. The ulna is followed by the humerus (13.07%), the radius (11.67%), the femur (10.78%), and the tibia at 9.35%.

The pelvis was also examined in greater detail. Fig. 5 shows a bar graph in which the portion of the pelvis under review is graphed along with the average error rate. Gathered data shows that isolating the Os Coxae yields an average error rate of 13.80%. By examining the pelvis as a whole, the average error rate drops to 5.62%.

The final point under review, again of heightened interest, was the methodology of determining biological sex versus the average error rate. Traditional methods rely more on the expertise of the observer, however, modern sex determining studies have utilized regression equations to make the distinction. Fig. 6 shows a bar graph where the two methodologies are compared based on their average error rate. The metric method, based on regression equations, was calculated to have an average error rate of 12.55%, while the non-metric, observation method had an average error rate of 10.46%.

Discussion

The analysis of the data shows that there is no significant relationship between the error in determining sex, and the number of traits used to make that determination. Therefore, it is the quality and not the quantity of specific traits that matter to forensic anthropologists when making the distinction between biologically male and female skeletons. Though it is beyond the scope of this paper, ideally, this work can contribute to the examination of which skeletal traits are most accurate when making a sex determination.

The skull was shown to be the least reliable in the determination of biological sex. However, by examining the skull as a whole, the average error rate was greatly reduced. Examining the skull in individual parts (mandible vs cranium) showed to have a higher average error rate.

Long bones were also examined for their efficacy in determining biological sex. Long bones had an intermediate average error rate. Of the long bones, the tibia showed the lowest average error rate, while the ulna showed the highest. Long bones have been used in court to determine sex in the absence of other, more traditional, portions of the skeleton. Understanding the possible hierarchy of accuracy using the different long bones should be pursued, as whole skeletons are rare in a forensic setting. It is necessary to understand the possible information that can be gleaned from all sections of the skeleton, and, for investigative reasons, it is important to understand how accurate those analyses might be.

The scapula and clavicle have less precedence in court, however, much research is under way to understand the type of information that can be derived from these bones. Studies determining their use in determining biological sex are promising. The gathered data put their average error rate above 10%, which can be problematic from an investigative standpoint. However, these numbers are affected by the limited sample size afforded in this paper. Studies of this nature are still in their infancy and could benefit from duplicate studies with different populations.

Out of the whole skeleton, the pelvis was shown to be the most accurate region in determining sex. However, full skeletal remains are a luxury very rarely enjoyed by anthropologists, and the pelvis may be fragmented, or entirely missing in a forensic context. In that case, other portions of the skeleton could be of use.

Based on the gathered data, all portions of the skeleton besides the pelvis had an average error rate greater than 10% which is usually considered unreliable in a forensic setting. However, this is most likely due to the limited sample size, as the goal of the study was to identify possible trends rather than definitive rates of error.

While modern times have been leaning more into the use of regression models to make certain determinations about skeletal remains, the gathered data shows that regression models may be less accurate than the determinations from experienced anthropologists. However, this finding is particularly limited by the sample size. There were significantly more studies utilizing regression models, versus those using visual analysis from experts.

All findings discussed are subject to the limitations of statistical analysis. The sample size of studies examined is cut down by the restrictions explained in the methods section. The studies picked were conducted during or after the year 2000. The selected studies also focused on modern populations. Importantly, all studies also focused on determining biological sex based on skeletal characteristics that have a precedent of admissibility in a United States court. Therefore, the sample size is a limiting factor is generalizing the findings discussed. Analysis of more specific variables are even further limited by sample size.

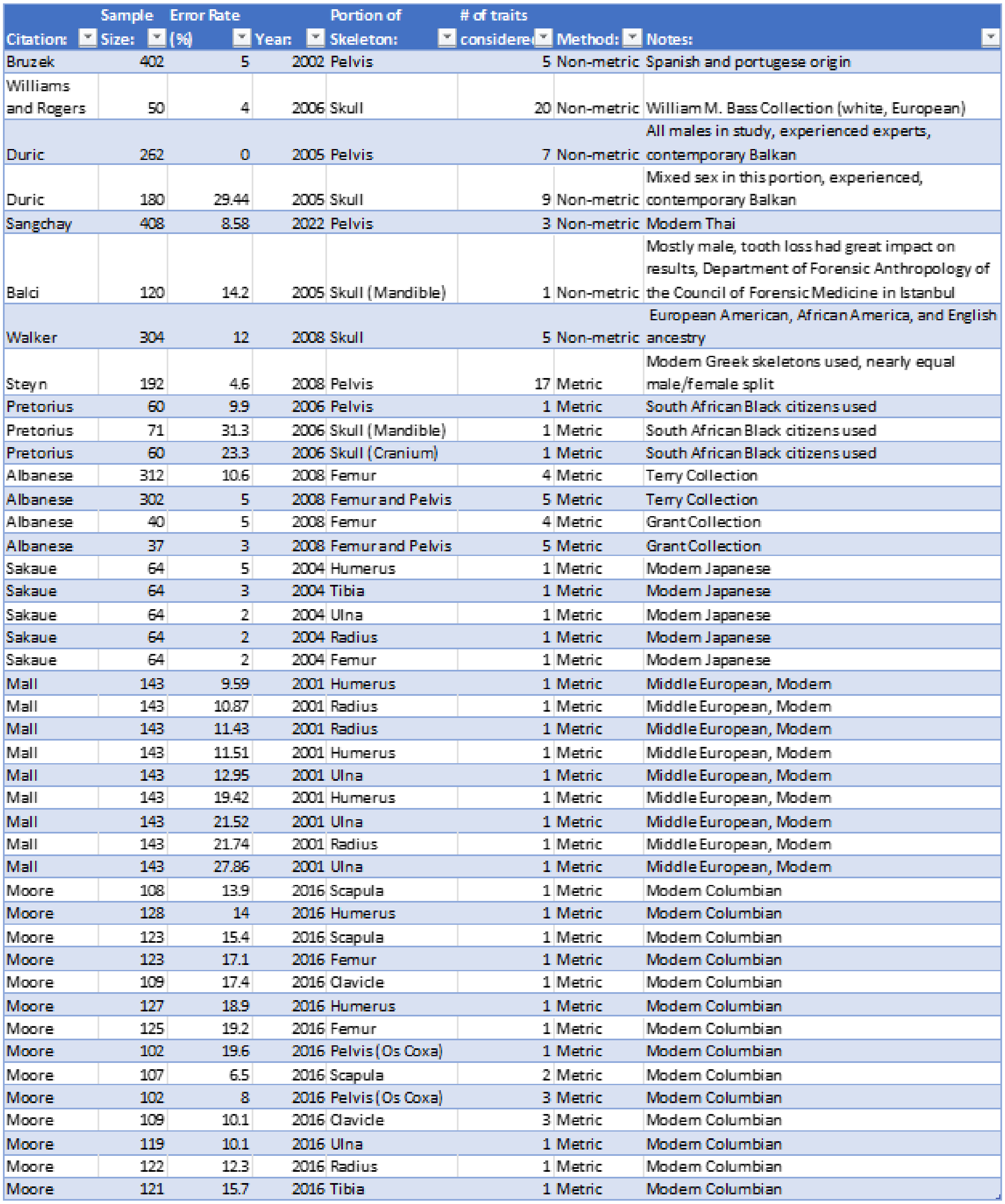
U.S. based studies with diverse representative populations could provide further insight into U.S. specific strengths and weaknesses in determining biological sex. Studies examining the qualities of different traits should also be conducted. Future emphasis should be on evaluating the quality of the markers used for biological sex, rather than increasing the number of traits that can be used for determining sex. Finally, most of the studies in the data set had the luxury of a comparative population. In a forensic context, this is rarely the case. Most determinations are made in a vacuum. So, studies testing the error rates on a single set of human remains, or ones that do not allow a comparative sample, could give a better indication of errors as they may appear in a forensic case today.

The study was conducted to identify possible trends in how specific variables affect the accuracy and precision of determining biological sex. The gathered data shows that the most accurate portion of the skeleton for determining sex is an intact pelvis. However, in the absence of a pelvis, long bones, and the scapula and clavicle, and the skull have traits that can differentiate the sexes. The emphasis should be on using high quality traits.

The research conducted here may potentially shine light on the most and least accurate means in determining biological sex for all involved in the carrying out of justice.

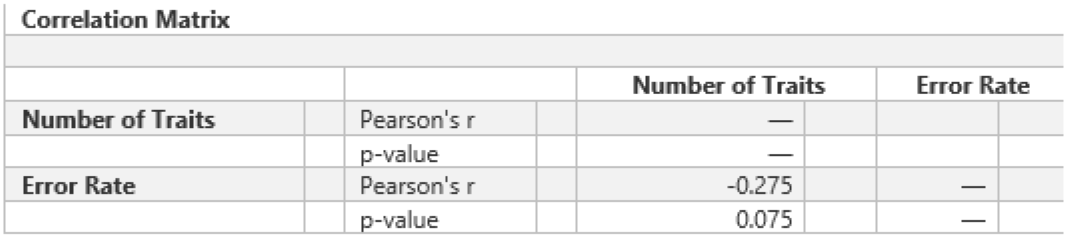
Forensic science has an everyday impact on the lives of people. It is the responsibility of forensic scientists to continuously vet their methods and upkeep rigorous standards of scientific quality. The overhaul of forensic science in the late 2000s largely kept forensic anthropology out of the picture, so methods in forensic anthropology are due for routine inspection of their efficacy.

Tables



Citation:	Sample Size:	Error Rate (%)	Year:	Portion of Skeleton:	# of traits considere	Method:	Notes:
Bruzek	402	5	2002	Pelvis	5	Non-metric	Spanish and portugese origin
Williams and Rogers	50	4	2006	Skull	20	Non-metric	William M. Bass Collection (white, European)
Duric	262	0	2005	Pelvis	7	Non-metric	All males in study, experienced experts, contemporary Balkan
Duric	180	29.44	2005	Skull	9	Non-metric	Mixed sex in this portion, experienced, contemporary Balkan
Sangchay	408	8.58	2022	Pelvis	3	Non-metric	Modern Thai
Balci	120	14.2	2005	Skull (Mandible)	1	Non-metric	Mostly male, tooth loss had great impact on results, Department of Forensic Anthropology of the Council of Forensic Medicine in Istanbul
Walker	304	12	2008	Skull	5	Non-metric	European American, African America, and English ancestry
Steyn	192	4.6	2008	Pelvis	17	Metric	Modern Greek skeletons used, nearly equal male/female split
Pretorius	60	9.9	2006	Pelvis	1	Metric	South African Black citizens used
Pretorius	71	31.3	2006	Skull (Mandible)	1	Metric	South African Black citizens used
Pretorius	60	23.3	2006	Skull (Cranium)	1	Metric	South African Black citizens used
Albanese	312	10.6	2008	Femur	4	Metric	Terry Collection
Albanese	302	5	2008	Femur and Pelvis	5	Metric	Terry Collection
Albanese	40	5	2008	Femur	4	Metric	Grant Collection
Albanese	37	3	2008	Femur and Pelvis	5	Metric	Grant Collection
Sakaue	64	5	2004	Humerus	1	Metric	Modern Japanese
Sakaue	64	3	2004	Tibia	1	Metric	Modern Japanese
Sakaue	64	2	2004	Ulna	1	Metric	Modern Japanese
Sakaue	64	2	2004	Radius	1	Metric	Modern Japanese
Sakaue	64	2	2004	Femur	1	Metric	Modern Japanese
Mall	143	9.59	2001	Humerus	1	Metric	Middle European, Modern
Mall	143	10.87	2001	Radius	1	Metric	Middle European, Modern
Mall	143	11.43	2001	Radius	1	Metric	Middle European, Modern
Mall	143	11.51	2001	Humerus	1	Metric	Middle European, Modern
Mall	143	12.95	2001	Ulna	1	Metric	Middle European, Modern
Mall	143	19.42	2001	Humerus	1	Metric	Middle European, Modern
Mall	143	21.52	2001	Ulna	1	Metric	Middle European, Modern
Mall	143	21.74	2001	Radius	1	Metric	Middle European, Modern
Mall	143	27.86	2001	Ulna	1	Metric	Middle European, Modern
Moore	108	13.9	2016	Scapula	1	Metric	Modern Columbian
Moore	128	14	2016	Humerus	1	Metric	Modern Columbian
Moore	123	15.4	2016	Scapula	1	Metric	Modern Columbian
Moore	123	17.1	2016	Femur	1	Metric	Modern Columbian
Moore	109	17.4	2016	Clavicle	1	Metric	Modern Columbian
Moore	127	18.9	2016	Humerus	1	Metric	Modern Columbian
Moore	125	19.2	2016	Femur	1	Metric	Modern Columbian
Moore	102	19.6	2016	Pelvis (Os Coxa)	1	Metric	Modern Columbian
Moore	107	6.5	2016	Scapula	2	Metric	Modern Columbian
Moore	102	8	2016	Pelvis (Os Coxa)	3	Metric	Modern Columbian
Moore	109	10.1	2016	Clavicle	3	Metric	Modern Columbian
Moore	119	10.1	2016	Ulna	1	Metric	Modern Columbian
Moore	122	12.3	2016	Radius	1	Metric	Modern Columbian
Moore	121	15.7	2016	Tibia	1	Metric	Modern Columbian

Table 1: A table outlining all of the data gathered for the analysis. The citation is listed, as well as the sample size, calculated error rate (in percentage), the year of the study, portion of skeleton under review, number of traits being considered, method, and notes on the population being studied.



Correlation Matrix

		Number of Traits	Error Rate
Number of Traits	Pearson's r	—	
	p-value	—	
Error Rate	Pearson's r	-0.275	—
	p-value	0.075	—

Table 2: A correlation matrix was run with the number of traits under consideration as the independent variable and the associated error rate as the dependent variable. The p-value was calculated as 0.075, which is not flagged as significant, as 0.075 is greater than 0.05.

Figures

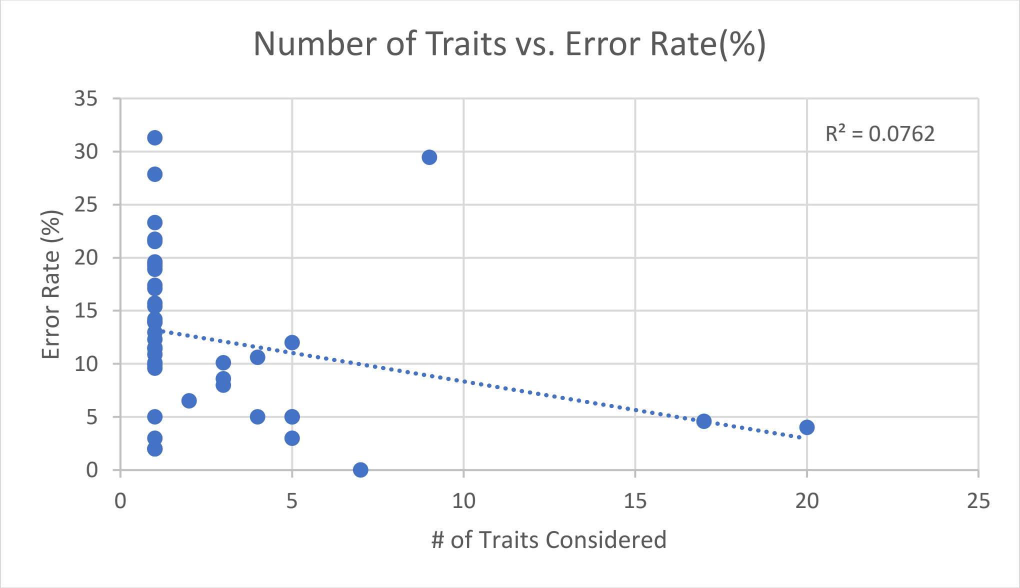


Fig. 1: Scatter plot of the number of traits being considered vs the error rate in determining biological sex. Both data points are pulled from Table. 1. Number of traits is the independent variable, while error rate is the dependent variable. A Pearson's correlation coefficient was calculated to be 0.0762, showing a very weak association between the two points.

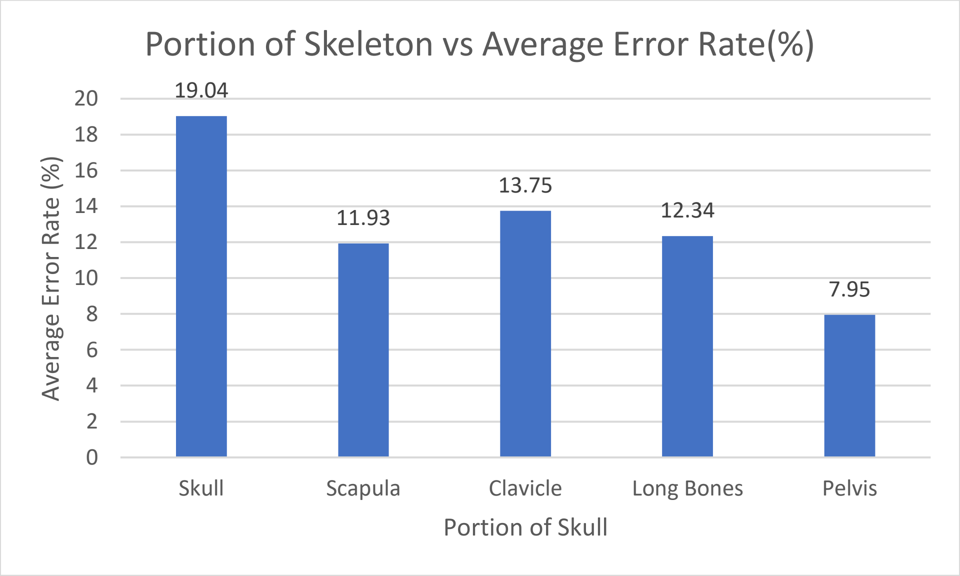


Fig. 2: Bar graph showing the portion of skeleton under review and the calculated average error rate. All data is pulled from Table. 1. Results show a 19.04 average error percentage for the skull; an 11.93 average error rate for the scapula; a 13.75 average error rate for the clavicle; a 12.34 average error rate for the long bones; and a 7.95 average error rate for the pelvis.

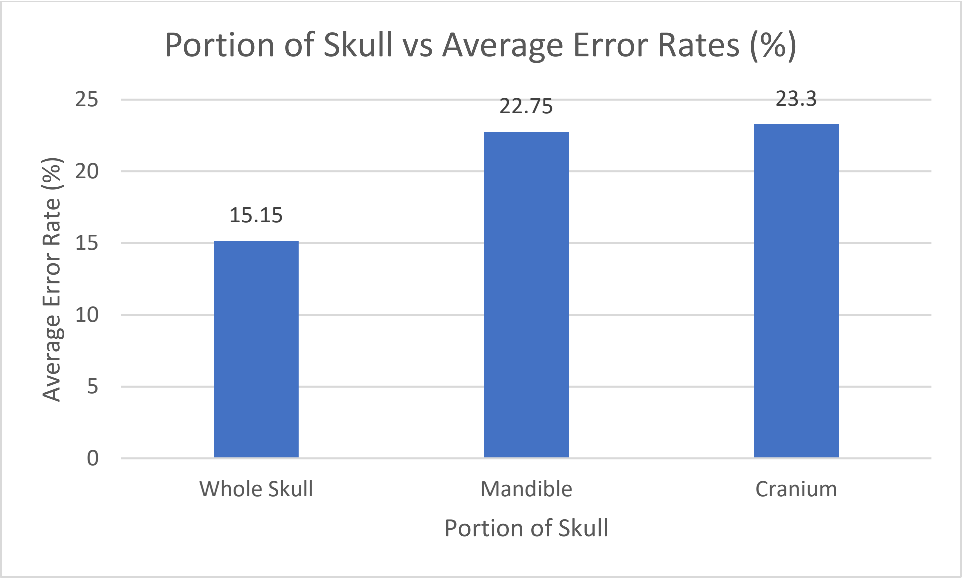


Fig. 3: A bar graph showing the portion of the skull under review vs the average error rate. All data is pulled from Table. 1. The whole skull shows an average error rate of 15.15. The mandible shows an average error rate of 22.75. The cranium shows an average error rate of 23.3.

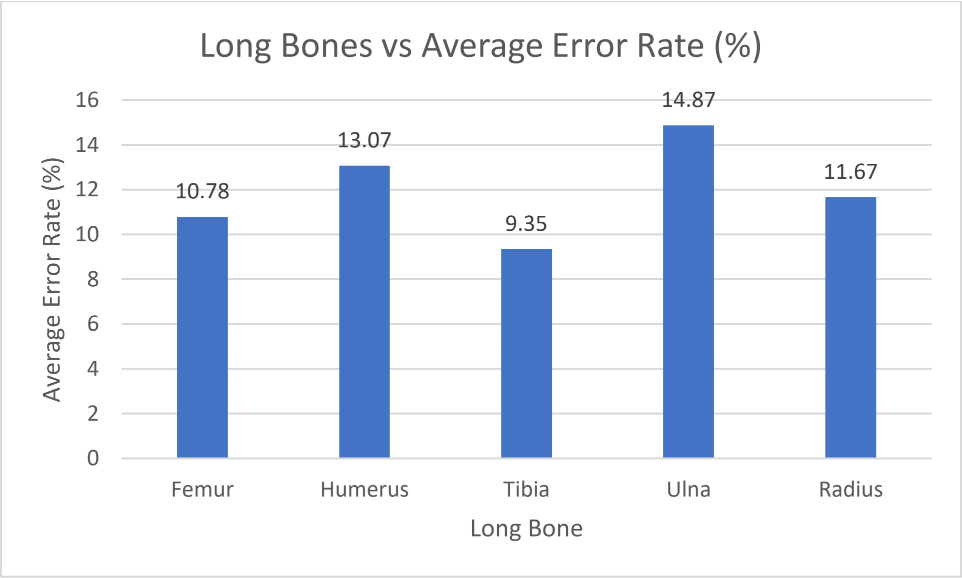


Fig. 4: A bar graph showing the average error rate for each of the long bones. All data is pulled from Table.1. The femur shows an average error rate of 10.78; the humerus, an average error rate of 13.07; the tibia, an average error rate of 9.35; the ulna, an average error rate of 14.87; and the radius, an average error rate of 11.67.

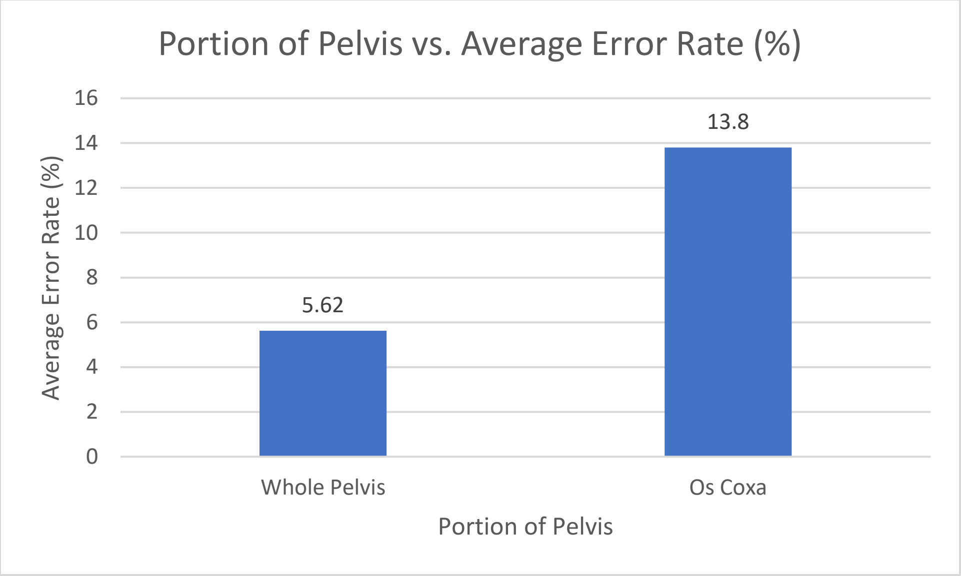


Fig. 5: A bar graph showing the portion of the pelvis being reviewed vs the average error rate. All data is pulled from Table 1. The whole pelvis shows an averge error rate of 5.62. When isolating the Os Coxa, the average error rate is 13.8.

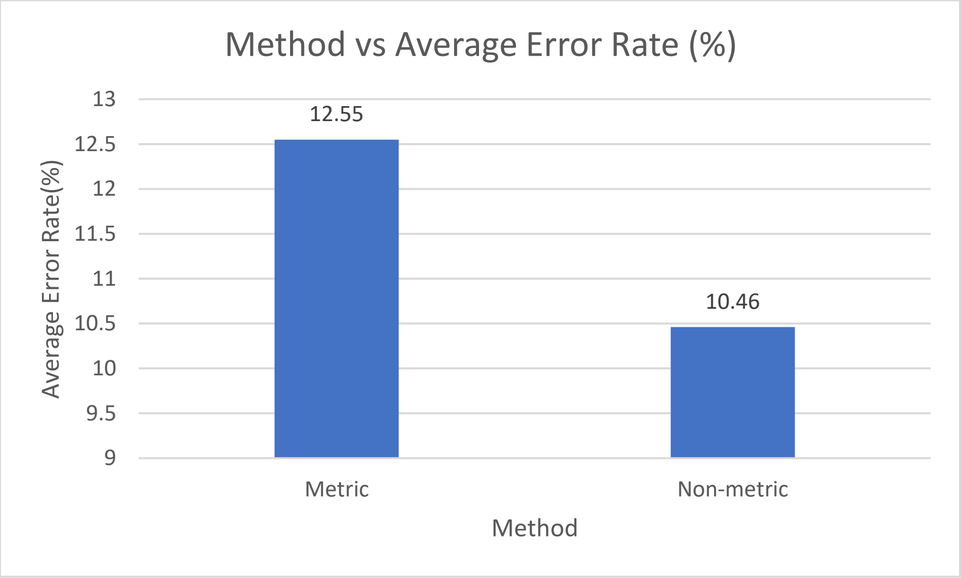


Fig. 6: A bar graph showing the methods of determination vs the average error rate. The methods are divided into metric and non-metric methodologies. The average error rate of the metric method was calculated to be 12.55, and the average error rate of the non-metric method was 10.46.

References

1. Medical Examiners And Coroners' Offices, 2004. Matthew J. Hickman, Ph.D., Kristen A. Hughes, M.P.A., *Bureau of Justice Statistics*, Kevin J. Strom, Ph.D., Jeri D. Ropero-Miller, Ph.D., DABFT, *RTI International.*
2. Innocence Staff (2019). Ten Years Later: The Lasting Impact of the 2009 NAS Report. Retrieved from https://innocenceproject.org/lasting-impact-of-2009-nas-report/.
3. National Research Council (U.S.). (2009). Strengthening forensic science in the United States: a path forward. National Academies Press.
4. Langley, N. R., Meadows Jantz, L., McNulty, S., Maijanen, H., Ousley, S. D., & Jantz, R. L. (2018). Error quantification of osteometric data in forensic anthropology. *Forensic Science International*, *287*, 183–189. https://doi.org/10.1016/j.forsciint.2018.04.004.
5. Valsecchi, A., Irurita Olivares, J. & Mesejo, P. Age estimation in forensic anthropology: methodological considerations about the validation studies of prediction models. *Int J Legal Med* **133**, 1915–1924 (2019). https://doi.org/10.1007/s00414-019-02064-7.
6. Wilson, R. J., Herrmann, N. P., & Jantz, L. M. (2010). Evaluation of stature estimation from the database for forensic anthropology. *Journal of Forensic Sciences*, *55*(3), 684–689. https://doi.org/10.1111/j.1556-4029.2010.01343.x
7. Curate, F. (2022). The Estimation of Sex of Human Skeletal Remains in the Portuguese Identified Collections: History and Prospects. *Forensic Sciences*, *2*(1), 272–286. https://doi.org/10.3390/forensicsci2010021
8. Spradley, M. K., Jantz, R. L., Robinson, A., & Peccerelli, F. (2008). Demographic change and forensic identification: Problems in metric identification of hispanic skeletons. *Journal of Forensic Sciences*, *53*(1), 21–28. https://doi.org/10.1111/j.1556-4029.2007.00614.x
9. Gualdi-Russo, E. (2007). Sex determination from the talus and calcaneus measurements. *Forensic Science International*, *171*(2–3), 151–156. https://doi.org/10.1016/j.forsciint.2006.10.014
10. Savag Giileq, E., Ya, M., & IFanb, sar. (1994). Forensic anthropology in Turkey. In *Forensic Science Forensic Science International* (Vol. 66).
11. Colman, K. L., Janssen, M. C. L., Stull, K. E., van Rijn, R. R., Oostra, R. J., de Boer, H. H., & van der Merwe, A. E. (2018). Dutch population specific sex estimation formulae using the proximal femur. *Forensic Science International*, *286*, 268.e1-268.e8. https://doi.org/10.1016/j.forsciint.2017.12.029
12. Kranioti, E. F., & Michalodimitrakis, M. (2009). Sexual dimorphism of the humerus in contemporary cretans - A population-specific study and a review of the literature. *Journal of Forensic Sciences*, *54*(5), 996–1000. https://doi.org/10.1111/j.1556-4029.2009.01103.x
13. Sanabria-Medina, C., González-Colmenares, G., Restrepo, H. O., & Rodríguez, J. M. G. (2016). A contemporary Colombian skeletal reference collection: A resource for the

development of population specific standards. *Forensic Science International*, *266*, 577.e1-577.e4. https://doi.org/10.1016/j.forsciint.2016.06.020

14. Bruzek, J. (2002). A method for visual determination of sex, using the human hip bone. *American Journal of Physical Anthropology*, *117*(2), 157–168. https://doi.org/10.1002/ajpa.10012
15. Williams, B. A., & Rogers, T. L. (2006). Evaluating the accuracy and precision of cranial morphological traits for sex determination. *Journal of Forensic Sciences*, *51*(4), 729–735. https://doi.org/10.1111/j.1556-4029.2006.00177.x
16. Durić, M., Rakočević, Z., & Donić, D. (2005). The reliability of sex determination of skeletons from forensic context in the Balkans. *Forensic Science International*, *147*(2-3 SPEC.ISS.), 159–164. https://doi.org/10.1016/j.forsciint.2004.09.111
17. Sangchay, N., Dzetkuličová, V., Zuppello, M., & Chetsawang, J. (2022). Consideration of Accuracy and Observational Error Analysis in Pelvic Sex Assessment: A Study in a Thai Cadaveric Human Population. *Siriraj Medical Journal*, *74*(5), 330–339. https://doi.org/10.33192/Smj.2022.40
18. Balci, Y., Yavuz, M. F., & Cağdir, S. (2005). Predictive accuracy of sexing the mandible by ramus flexure. *HOMO- Journal of Comparative Human Biology*, *55*(3), 229– 237. https://doi.org/10.1016/j.jchb.2004.07.006
19. Walker, P. L. (2008). Sexing skulls using discriminant function analysis of visually assessed traits. *American Journal of Physical Anthropology*, *136*(1), 39–50. https://doi.org/10.1002/ajpa.20776
20. Steyn, M., & Işcan, M. Y. (2008). Metric sex determination from the pelvis in modern Greeks. *Forensic Science International*, *179*(1), 86.e1-86.e6. https://doi.org/10.1016/j.forsciint.2008.04.022
21. Pretorius, E., Steyn, M., & Scholtz, Y. (2006). Investigation into the usability of geometric morphometric analysis in assessment of sexual dimorphism. *American Journal of Physical Anthropology*, *129*(1), 64–70. https://doi.org/10.1002/ajpa.20251
22. Albanese, J., Eklics, G., & Tuck, A. (2008). A metric method for sex determination using the proximal femur and fragmentary hipbone. *Journal of Forensic Sciences*, *53*(6), 1283–1288. https://doi.org/10.1111/j.1556-4029.2008.00855.x
23. Sakaue, K. (2004). Sexual determination of long bones in recent Japanese. *Anthropological Science*, *112*(1), 75–81. https://doi.org/10.1537/ase.00067
24. Mall, G., Hubig, M., Èttner, A. B., Kuznik, J., Penning, R., & Graw, M. (n.d.). *Sex determination and estimation of stature from the longbones of the arm*

25. Moore, M. K., DiGangi, E. A., Niño Ruíz, F. P., Hidalgo Davila, O. J., & Sanabria Medina, C. (2016). Metric sex estimation from the postcranial skeleton for the Colombian population. *Forensic Science International*, *262*, 286.e1-286.e8. https://doi.org/10.1016/j.forsciint.2016.02.018